I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14115

Kinks as Goals: Accelerating Commissions and the Performance of Sales Teams

Peter Kuhn
Lizi Yu

FEBRUARY 2021

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14115

Kinks as Goals: Accelerating Commissions and the Performance of Sales Teams

Peter Kuhn
University of California, Santa Barbara and IZA

Lizi Yu
University of Queensland

FEBRUARY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Kinks as Goals: Accelerating Commissions and the Performance of Sales Teams*

We study the performance of small retail sales teams facing an incentive scheme that includes both a lump sum bonus and multiple accelerators (kinks where the piece rate jumps upward). Consistent with standard labor supply models, we find that the presence of an attainable bonus or kink on a work-day raises mean sales, and that sales are highly bunched at the bonus; inconsistent with those models we find that teams bunch at the kinks instead of avoiding them. Teams' responses to the kinks are consistent with models in which the kinks are perceived as symbolic rewards, and inconsistent with reference point models where kinks induce loss aversion.

JEL Classification: J33, M12

Keywords: teams, bonuses, bunching, accelerators, symbolic rewards, loss aversion

Corresponding author:
Peter Kuhn
Department of Economics
University of California, Santa Barbara
Santa Barbara, CA 93106
USA
E-mail: peter.kuhn@ucsb.edu

* We thank Youssef Benzarti, Clement de Chaisemartin, Alex Rees-Jones, Alisa Tazhitdinova, Gonzalo Vasquez, and participants at the 2019 Trans-Pacific Labor Seminar (Tokyo) for helpful comments.

1 Introduction

Putting an accelerator or accelerators into the commission plan is the most common method applied. Put simply, the rate of commission paid is accelerated for every dollar earned above the set sales target. This encourages the salesperson to not only meet their sales targets, but to 'smash their sales targets'! (MyFirstSalesJob, 2020)

Despite theoretical arguments that non-linear incentive pay schemes distort agents' behavior (Holmstrom and Milgrom, 1987), a wide variety of non-linear schemes are used in workplaces.[1] Empirical studies of non-linear pay schemes, however, have focused heavily on one particular type of non-linearity: lump-sum bonuses for attaining a target level of output.[2] While adding a bonus to a compensation plan can raise performance (Freeman et al., 2019), a number of studies also confirm the types of distortions predicted by Holmstrom and Milgrom, which include mis-pricing of the product (Larkin, 2014; Owan et al., 2015), inefficient personnel retention decisions and distortions of subordinates' incentives (Benson, 2015), and manipulating sales timing in ways that harm a company's reputation.

The goal of this paper is to study a second common form of non-linearity in incentive pay plans, namely *accelerators*.[3] Accelerators are performance thresholds at which the *slope* of the pay schedule jumps upward, and appear to have received much less attention.[4] In principle, accelerators could affect workers' effort and performance through at least three channels. First, accelerators are sometimes combined with a lump sum reward. In these cases, both the discrete reward and the change in slope will affect performance– the discrete reward incentivizing workers to attain the threshold and the accelerator incentivizing them to exceed it. Second, accelerators divide the set of feasible performance levels into salient, disjoint ranges, which frequently have labels attached

[1]Non-linear pricing is also used in retail sales. To the best of our knowledge, most of this work focuses either on quantity discounts associated with different package sizes (Cohen, 2008), or on rising marginal prices for products like electricity and cell phone service (Lambrecht et al., 2007). Neither of these correspond directly to the lump-sum bonuses or accelerators that are commonly used in incentive pay; in the first case because only a finite number of package sizes are available, in the second because quantity premia would correspond to *decelerators* in incentive pay schemes.

[2]Douthit (1976) and Joseph and Kalwani (1998) respectively report that 80 and 72 percent of firms use bonuses to incentivize their salespeople. Oyer (2000) reports that 89 percent use sales quotas (which could apply to bonuses or accelerators).

[3]O'Connell (1989) states that 53 percent of sales compensation plans have increasing commission rates; Parrinnello (2017) reports that over-quota accelerators are used in in most sales compensation plans. Convex pay schedules may also be attractive to employers for theoretical reasons. For example, Basu et al. (1985) show that convex pay schedules are efficient when a salesperson's (absolute) risk aversion declines with income, and Oyer (2000) shows that convex compensation plans yield higher profits than linear ones when workers face liability limitations (i.e. they cannot 'pay for jobs'). This is because convex schemes reduce the need to pay positive marginal compensation for low realized sales levels. Oyer's result parallels an earlier finding by Spence (1977) who argued that the profit-maximizing nonlinear pricing policy for a firm with monopoly power involves quantity discounts, which are the product market counterpart to accelerating commissions.

[4]The performance pay plan (PPP) studied by Lazear (2000) can be thought of as an accelerating commission scheme in which the commission rate is zero on sales below a certain level. Lazear did not study the distribution of workers' output levels around this kink, however, nor did he use such information to uncover the mechanisms explaining the accelerator's effects.

to them, such as “Tier1” and ”Tier 2”, or “Good”, “Excellent, and “Superior’. These ranges could function as reference points or symbolic rewards that affect worker behavior. Finally, the increasing marginal pay associated with accelerators directly encourages workers to exceed the threshold, but –at least in the absence of attached bonuses or ‘behavioral’ determinants of effort– actually incentivize workers to *avoid* output levels just around the threshold.

To understand how and why a typical accelerating commission schedule affects worker behavior, this paper studies the performance of small sales teams in 103 stores operated by a large clothing firm in China (hereafter “Firm A”) whose monthly pay schedule includes three thresholds. At the lowest of these (henceforth the *target*) the commission rate rises and a small lump-sum bonus (worth about 3.5 percent of mean monthly pay) is awarded. At the other two thresholds, (henceforth the *kinks*) only the commission rate rises.[5] The four team output intervals corresponding to these three thresholds are labeled as “Standard”, “Good”, “Excellent”, and “Superb” on a worker’s monthly pay stub. That said, no other monetary rewards, such as promotion decisions are linked to attaining these ranges. Indeed, because the bonus is team-based and the company does not collect information on individual worker performance, there is no way for an individual worker to single herself out for promotion by performing better than her teammates.[6] Nevertheless, workers may attach a symbolic value to attaining these ranges as a team.

We start our analysis by deriving the implications of the standard labor supply model for the distribution of team output around these two types of thresholds.[7] We then compare these predictions with empirical estimates of output distributions at both the monthly and daily levels, in the latter case applying a *difference in density differences* (DDenD) estimator to quasi-random variation in the presence and location of an accelerator on the last day of a sales month. Consistent with the standard labor supply model, we find strong bunching of output levels at the target (where a bonus is paid). Also consistent, we also find that the presence of an attainable target *or* kink on the final day of a month raises mean sales on that day. Inconsistent with the standard model –which predicts missing density in the sales distribution at concave kinks– we also find strong bunching at the two pure kinks.[8] This bunching is surprisingly precise –concentrating in the $100 bin that contains the kinks– and is robust to controlling for a rich set of cyclical effects and team heterogeneity. Finally, we assess the ability of two behavioral models –one in which the kink acts

[5]According to Firm A, this type of schedule –a target with a bonus attached, plus accelerators beyond that– is also common among its competitors in the region. Thus we are studying a system that has withstood the test of time and competition, not –as is common in field experiments– an unfamiliar policy in a particular firm.

[6]It is possible that team *leaders* whose teams consistently pass the kinks have a better chance of being hired into managerial positions, though such promotions are infrequent and the firm does not consider this an important effect.

[7]Throughout the paper we will model the behavior of sales teams as if they were individual workers. If free-rider effects mute the effects of incentives on performance, our estimates are likely to understate the effects of individual-level incentives of this type.

[8]Kleven (2016) distinguishes between convex kinks (which maintain the convexity of the budget set and induce bunching) and non-convex kinks, which should produce gaps. We use the term *concave* for the second type of kink. Kleven reports that existing research on concave kinks is rare, and has not found much evidence of missing mass there. Seibold (2021) is the only study we know of that, like us, finds bunching at concave kinks. Those kinks appear to be driven more by a lack of information about Germany’s retirement system than by the symbolic incentives we argue are at work in our context.

as a loss-aversion-inducing reference point, and one in which workers derive direct utility from meeting or exceeding the kink– to account for this unexpected result, and we argue that only the second, 'symbolic rewards' model is consistent with all the patterns in the data.[9]

In view of the well-documented gaming of bonus incentive schemes, it is noteworthy that several of the common unintended effects of non-linear pay are not relevant in our context. For example, mis-pricing is not possible because employees have almost no discretion on price. Workers' ability to manipulate the timing of sales is limited because all sales are instantly recorded electronically, and relationships with customers are mostly short-term. Hasty sales will not result in returned items because purchases made at physical stores are not refundable at Firm A, which is typical at physical stores in China.[10] At the same time, the sales target performs an additional role for the firm that we do not model formally in the paper: In a high worker turnover context where sales traffic differs widely between locations and exhibits high fluctuation, the target –which equals each store's sales in the same calendar month from the previous year– is a simple and effective way to communicate expected performance levels to workers.[11] Overall, our analysis suggests that (a) accelerators are effective in motivating workers, and (b) their effect is not limited to the direct financial conquences of increasing marginal rewards. Accelerators have an additional motivational effect that is tied to the direct utility value of attaining different ranges of the compensation schedule.[12]

In addition to the literature on non-linear reward schedules, our paper contributes to the literatures on symbolic rewards and goals. While a number of authors have studied symbolic rewards, much of the existing work focuses on *relative* symbolic rewards that induce a tournament among employees (Kosfeld and Neckermann, 2011; Ashraf et al., 2014). Awards of this type can have negative spillover effects on the workers who do not receive them, which can wash out their incentivizing effects on the winners.[13] In our case, the symbol is activated when the team (comprising all the employees of the company at a particular workplace) attains a pre-specified *absolute* level of performance. Since everyone in a worker's natural reference group either receives or does not receive the symbolic reward, our rewards should induce co-operation, not competition. Thus, an additional finding of our paper is that *team-based* symbolic rewards based on *absolute* performance can be motivating, without the negative spillover effects associated with awards for relative, individual performance.

[9] Specifically, while both models *can* generate bunching at a concave kink, the reference-dependence model can only do so if loss aversion is sufficiently strong relative to the rise in commission rates at the kink. Furthermore, if bunching occurs, the reference-dependence model predicts that introducing a kink reduces the density of output strictly above the kink, which is inconsistent with our results.

[10] It is of course still possible for unsatisfactory purchases to hurt the firm's reputation, but Firm A does not consider this to be a major issue of its pay scheme.

[11] While a target based on last year's sales raises the potential for ratchet effects, this is not a concern during our data period. Please see Section 2.2 for details.

[12] Gu and Yang (2010) introduce non-standard preferences to the nonlinear consumer pricing literature by letting the price per unit enter directly to transaction utility.

[13] Such spillovers have also been observed when awards are given for high *absolute* levels of performance, such as perfect attendance (Gubler et al., 2016).

The literature on goals studies situations like ours where the set of feasible performance levels is divided into intervals with potentially meaningful labels attached to them. In our reading, the existing work focuses mostly on individual performance targets that are chosen by workers themselves. Typically, these *commitment contracts* impose a monetary cost on workers who do not attain them (Royer et al., 2015; Kaur et al., 2015).[14] More closely related to our work, Agarwal et al. (2017) study exogenously-assigned goals for a prosocial activity–resource conservation– with no financial consequences.[15] Like us, they find that goals work best when they are neither too easy nor too hard, but they suggest that the goals' effects are explaiined by loss aversion.

Finally, our paper contributes to a growing applied literature devoted to estimating the amount of bunching, or 'excess density' in distributions of economic outcomes at theoretically relevant values. Early approaches to this question (Chetty et al., 2011; Saez, 2010; Kleven and Waseem, 2013) studied large samples of draws from what is assumed to be a single distribution, in which there is a fixed potential bunching point of interest. For example, one may have millions of observations on the level of pre-tax income, and one is interested in the extent to which taxpayers bunch at the maximum EITC entitlement. The authors then fit a high-dimensional polynomial to the entire frequency distribution –excluding a small region of interest– and use the implied smoothness assumptions to identify the excess mass in that region.

In our case, while we observe team sales on a large number of days, our models suggest that we should not see any bunching or gaps on most of those days (either because there is no threshold within the team's reach, or because reaching the threshold on that day is not a necessary condition for attaining it for the month). This gives us a large sample of *control days*. On the other days (which are a subset of the last days of the month), we expect bunching or gaps at particular sales levels that vary quasi-randomly from day to day, and from team to team. Our setting is complicated by the fact that other observable factors, such as day of week and store location, also affect the daily sales distribution and should be controlled for. As noted, we address this issue using a "difference in density differences" (DDenD) estimation approach that allows us to non-parametrically estimate the entire frequency distribution of daily team sales under a particular set of conditions.[16] While our approach is not optimal for all contexts, it allows us, for example, to measure the extent to which the presence of an attainable kink in the $1100-$1199 bin on a given day and store shifts the entire distribution of output, not just in the area around the bin itself. These features may prove useful in other studies of the effects of nonlinear incentive plans, including bonuses, reference points, and payday effects, among others.

[14] Dobronyi et al. (2019) study such goals when no monetary cost is attached, and find they have a zero effect.

[15] Our context is also similar to the conversion of numeric to letter-based academic grades (Oettinger, 2002), though letter grades have real 'lump sum' consequences for academic promotion. In addition, letter grading schemes frequently have a competitive, zero-sum feature that is absent in our context.

[16] More recently, Fortin et al. (2021) have used a probit-based *distribution regression* approach to estimate the effects of covariates on the entire *cdf* of wages. Our approach is more closely related to Cengiz et al. (2019), who directly estimate frequencies in all the bins of a wage distribution when a theoretically relevant quantity –in their case the minimum wage– falls into different bins in different times and places. Pierce et al. (2020) use a similar, but kernel-based, approach to estmate the effects of loss-framed performance incentives on sales distributions.

2 Background and Data

2.1 Retail Jobs and Firm A

Our data represent 103 retail stores operated by Firm A, a large manufacturer and retailer of men's clothing in China.[17] We study the performance of these stores during 2016, when the incentive scheme we analyze was in place.[18] Dropping observations in the month of store openings, closures or remodeling, we have 34,863 daily observations of team-level sales. The company does not maintain individual employee sales data, nor does it use them in setting pay.[19]

Firm A operates retail stores in two types of locations: department stores and shopping malls. A typical store includes a counter and a display of products, and is operated by a sales team of 2-7 employees. According to the frontline salespeople, the yield rate per customer, i.e. the fraction of people who walk into the store who actually make a purchase, is quite high for men's clothing retail, at between 50 and 90 percent.[20] Despite this high yield rate, employees and managers believe that sales amounts are quite sensitive to employee effort. In part this is because Firm A's salespeople play a more active role than many U.S. salespeople: With very limited display space, most items must be retrieved from storage to be tried on. Customers also place a high value on fit and speed of service, both of which are sensitive to employee effort.

As retail sales also depend on customer traffic, daily sales fluctuate quite dramatically over time, as shown in Figure 1. This figure plots the average daily sales per store on every calendar day in 2015 and 2016. The smaller cycles in the figure represent output within a week, with higher daily sales on weekends. The largest spikes are labelled, and correspond to major holidays when people shop for menswear heavily, such as Chinese New Year or Father's Day. Controlling for these large day-of-week and holiday effects will play a significant role in our econometric analysis. As shown in Table 1, the average daily sales output was $582 in 2016, but daily sales were both 'lumpy' and unpredictable: only 11.3 relatively high-value items were sold per day, and the standard deviation of daily sales ($1241) was over twice the mean and 4.5 times the median.

[17] Our sample excludes 46 stores whose workers were not directly hired and paid by Firm A. Since the teams' monthly targets, T, were given by their own sales in the previous year (2015), our sample also excludes stores that did not operate in 2015.

[18] Firm A's salespeople had experience with similar types of incentive schemes in previous years, but the details varied somewhat from year to year. More specifically, by 2016, Firm A had used a similar nonlinear schedule –containing bonus(es) and accelerator(s)– for at least five years, but the targets were either determined by store employees themselves (subject to management approval), or by regional or higher-level managers.

[19] See Section 2.2 for a more detailed description of the compensation schedule.

[20] As documented in the marketing literature (Venkatesh and Agarwal, 2006), male shoppers are more utilitarian and loyal customers who shop for a purpose.

2.2 The Compensation Schedule

The employees in our sample are paid monthly, including a base salary and a commission. In 2016, the average monthly compensation was \$545, around 40% of which was commission. Within a store, the base salary varies with employees' tenure, while commission rates are identical for all employees. Commissions are based on the *store's* total monthly performance; they are set lower in larger stores, in order to yield similar daily wages across teams of different sizes, as indicated in Table 1.

As illustrated in Figure 2, Firm A's nonlinear incentive plan in 2016 includes a target at which a bonus is awarded and the commission rate increases, plus two pure accelerators where only the commission rate rises. In particular, a store's target T is defined as its sales in the same calendar month in the previous year (2015).[21] If a store attains T, each team member receives a lump sum bonus of \$15.63 (100 CNY), and is exempt from a \$3.13 (20 CNY) penalty for not meeting the target, resulting in a jump equivalent to \$18.76 (120 CNY). This is worth 3.5% of employees' average monthly compensation. At the target T, the commission rate accelerates to 1.5 times the baseline rate. Next, there are two pure kinks in the schedule: The first accelerator is at $1.3T$, where the commission rate rises to 1.8 times the baseline rate, and the second is at $1.6T$, beyond which twice the baseline commission rate is paid. At the beginning of the year, regional managers communicated this annual sales plan to the store managers.[22]

In interviews, management described three main motivations for their use of this pay scheme. First, they believe that the target T is a simple and effective way to communicate their expected performance level to each store, in a context where customer traffic varies dramatically across stores and across months of the year (Ockenfels et al., 2015). Second, as argued in Kuhn and Yu (2021), Firm A's use of *team*-based incentives is part of an HR system that effectively delegates many key decisions, including recruiting and employee discipline, to the team members.[23] Third, the firm's use of accelerating piece rates without bonuses at higher performance levels ($1.3T$ and $1.6T$) is based mostly on a stated desire to be consistent with other local retailers and retain high-performing salespeople. Overall, the target, kink positions, bonus amount and accelerators are

[21] With forward-looking employees, a sales target mechanically linked to past output could lead to strategic output restriction, or ratchet effects (Charness et al., 2011), which could make the sales target an endogenous result of worker behavior. This is not a concern in our data because (a) turnover at Firm A is high, (b) 2016 was the first year in which sales targets were linked to the previous year's performance, and (c) the decision to link 2016 targets to 2015 sales was announced after 2015 sales were realized.

[22] Within the same year, sales targets will not be adjusted even if a store experiences employee turnover. If a store fails to hire a replacement after some period, store managers will inform the regional managers, and typically commission rates will be adjusted to compensate remaining employees' extra efforts, while sales targets and kinks remain unchanged.

[23] Despite the theoretical possibility of free-rider problems, team-based incentives are used quite widely, including Continental Airlines (Knez and Simester, 2001), Microchip (Adamson et al., 2014), U.S. steel minimills (Boning et al., 2007), and apparel manufacturing (Berg et al., 1996; Hamilton et al., 2003). Examples from retail include Wal-Mart's profit-sharing plan, team-based bonuses in German retail establishments (Friebel et al., 2017), and tip pooling in restaurants (Scudder, 2017). Lawler and Mohrman (2003) report that the share of Fortune 1000 companies using work-group or team incentives for more than a fifth of their workers more than doubled, from 21 to 51 percent, between 1990 and 2002.

within the range used by the Firm A's labor market competitors.

According to Firm A, except for the bonus at T, no additional monetary rewards are attached to the nonlinear incentive plans. However, the presentation of pay information in workers' monthly pay slips may attach symbolic rewards to this incentive scheme. Figure 3 presents a template of the pay slip distributed to salespeople. Specifically, the commissions are calculated and recorded separately for the four sales ranges that correspond to the available commission rates (i.e. $1\times$, $1.5\times$, $1.8\times$, or $2\times$ the baseline). On the pay slip, these ranges are labelled as "Standard", "Good", "Excellent", and "Superb", respectively. Moreover, at the bottom of the pay slip, there is an extra line called "Assessment", which simply reports the highest of these categories attained by the team. We do not know precisely when this practice started, but salespeople recall instances from as early as 2014.

3 Theory– Predicted Output Distributions

In this section we compare the predictions of three simple models for the empirical distributions of output around the two types of thresholds in Firm A's pay schedule– the target (T) and the kinks ($1.3T$ and $1.6T$). The *standard model* assumes that workers derive utility only from money and leisure, and have quasi-linear utility in money. Following Saez (2010) all our models assume that workers differ according to a single cost-of-effort parameter that is strictly positive and continuous on $(0, \infty)$. In the *symbolic rewards (SR)* model, in addition to the standard deteriminants of income and effort, teams perceive attaining the kink as a (lump sum) psychic reward that directly raises their utility. Finally, we consider a *reference point (RP)* model in which a kink creates a reference point, and induces loss aversion relative to it. Appendix A provides additional detail on the structure of the three models and derives the implications summarized below.

Turning first to the models' predictions for the two 'pure' kinks ($1.3T$ and $1.6T$), the standard model predicts a gap, or missing density, in the distribution of output in an interval that (strictly) includes the kink, i.e both at the kink and on both sides of it. This result is well known in the literature on bunching estimators (Kleven, 2016). The symbolic rewards model, on the other hand, predicts missing density to the left of the kink, *bunching* at the kink, and excess density to the right of the kink. Intuitively, the symbolic reward is equivalent to a bonus that pulls some people up to the threshold, while the increase in slope pulls some people past it. Finally, the predictions of the reference-point model depend on whether the loss aversion around the kink is strong or weak relative to the change in the commission rate. If loss aversion is not strong enough to outweigh the effects of the rising cash incentives at the kink, the model's predictions are the same as the standard model– a gap in the output distribution around the kink. If loss aversion outweighs the acceleration in financial rewards, we should see bunching at the kink, but *missing* mass to its right.

Appendix A also derives the predictions of these three models for the distribution of sales output around the target, T. Interestingly, in three of the four cases these predictions are qualititatively the same as the symbolic rewards model for the *kinks*. The intuition is that the cash bonus has the same effect on the output distribution as a lump-sum symbolic reward. Finally, in the remaining case –strong loss aversion– the predictions for the target are the same as the symbolic rewards model's predictions for the kinks. For easy reference, all these predictions are summarized in Table 2, which also highlights the output patterns that we find –quite robustly– in our empirical analysis. Among the models considered, only the symbolic rewards model can account for the empirical patterns around both the target and the kinks.

4 Descriptive Evidence

This section establishes three facts that motivate and guide our main empirical analysis. First, using *monthly* sales data, we establish that there is statistically significant bunching at both the target, T, and the first kink, $1.3T$. While we do not see bunching at the second kink, $1.6T$, we do not see the gap that is predicted by the standard model either. Second, using *daily* sales data, we show that there is bunching at all three thresholds (T, $1.3T$, and $1.6T$) when the threshold is *within reach* (i.e. within three days' worth of sales), but only on the last days of a calendar month.[24] Since attaining a monthly threshold is only 'at risk' on the last day of the month, this suggests that Firm A's work teams place some value on attaining salient ranges of the monthly output distribution.[25]

Third, we show that there are no statistically significant differences between the daily output distribution on last days of the month when no threshold is within reach and the distribution on *all* non-last days of the month. Both these distributions, however are distinct from last days when a threshold is within reach. This suggests that we can use both these types of days as controls for last days on which a threshold is within reach. Thus, our main estimation strategy will compare the distribution of a store's output on 'treatment' days –i.e. last days on which a threshold is within reach– to all other other days of a month in that store.[26]

4.1 Bunching in the Monthly Sales Distribution

As a first look at sales patterns in Firm A's stores, Figure 4(a) plots the distribution of sales across 1,143 store-month observations in 2016, where sales are measured relative to the team's target T (i.e. its same-month sales in 2015). For every store, observations are grouped into bins of width $0.1T$ that are aligned with the target ($0.8T$ to $0.9T$, $0.9T$ to T, etc); consistent with Firm

[24] In our baseline specification, 'days of sales' refer to a store's own mean daily sales during the previous days of the current month. Figure 5, as well as our main estimates, are robust to other definitions of being within reach– see Section 4.3.

[25] By 'at risk' we mean that reaching the threshold today is a necessary condition for attaining it in the current month.

[26] Appendix C shows that our main results are unchanged when we use a number of other comparison groups, for example excluding the last seven days of the month. Note also that in our main analysis, the treatment status of a last day will be specific to a particular threshold. For example, $1.3T$ might within reach but not T (because the team has already exceeded it), or $1.6T$ (because it is too far away). In fact, it is almost never the case that more than one threshold is within reach on the same day. This is because the thresholds are about 0.3*30 = 9 days' worth of sales apart, but a threshold must be within three days' of sales to be within reach.

A's target-setting formula, sales teams fail to reach their monthly target in 50% of all store-month observations. Consistent with the standard model described in Section 3, we see clear bunching at T; this is, in fact where the bunching is most striking. Contrary to the standard model, however, we observe some bunching at $1.3T$. In contrast, there is no visual evidence of bunching, nor of the predicted gap, at the last kink point $1.6T$. Aside from these two mass points, the histogram elsewhere exhibits a smooth right-skewed bell shape.[27]

To quantify these visual impressions, Figure 4(b) applies the approach developed by Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013) to estimate the amount of excess density around T, $1.3T$, and $1.6T$. Using the number of observations that fall within each bin, we fit a high-degree polynomial (8th degree), excluding data in those three bins.[28] Then, using the estimated coefficients from the polynomial, we extrapolate a fitted distribution to the excluded points to estimate the counterfactual distribution.[29] The estimated bunching using this approach is 0.3030 at T, which is significant at 1% using the bootstrapped standard error. This implies that the excess mass around the target is 30% of the average height of the counterfactual bin count. The estimated bunching at $1.3T$ is 0.2194, which is also significant at 1%. In contrast, at $1.6T$, we find negative bunching of -0.1002, but the estimate is not significant.[30]

While Figure 4(b) provides preliminary evidence of bunching around the target and the first kink, our ability to make sharp statements using this approach is limited by the fact that we have only 1,143 observations of monthly store performance, compared for example to the 11.6 million and 4 million observations used by Chetty et al. (2011) and Kleven and Waseem (2013) respectively. These limitations motivate our focus on daily sales data, which dramatically raises our sample size, and –as we shall argue– yields two advantages that are not available in Figure 4(b)'s approach: (a) it generates quasi-random variation in the presence *and* location of a threshold in the current day's reward schedule, and (b) it eliminates the need for a smoothness assumption to identify the amount of excess density around the target and kinks.

4.2 Bunching in the Daily Sales Distribution

To illustrate the relationship between teams' output thresholds and the *daily* distribution of sales, we start by measuring each team's 'inherited' output at the start of every day, t, as the sum

[27] The observations in the tails are mostly generated by the fact that the company's targeting formula fails to account for holidays or sales promotions that occur in different months in different years. For example, the Chinese New Year arrived in February 2015 but in January 2016; as a result the teams' 2016 sales targets were very easy in January, and quite unrealistic in February.

[28] In a typical application of this method, a narrow range of bins near the hypothesised bunching location will also be excluded, to allow the excess mass to diffuse around these points. We do not exclude neighboring bins for two main reasons. First, we would have to exclude all bins from $0.9T$ to $1.7T$, as the three kinks in our context are close to each other, and that would lose a major part of the distribution. Second, the bins used in our analysis already allow for considerable noise around the kink, because each bin of $0.1T$ is equivalent to three average days' output.

[29] Details of this procedure are provided in Appendix B.

[30] Interestingly, this figure also shows evidence of bunching at $2T$, which supports the possibility that even numbers also serve as symbolic rewards, or reference points. This phenomenon is also found in Benson (2015) and Allen et al. (2017).

of sales from the first day of the month up to and including day $t-1$. Then, separately for each of the three thresholds, we restrict our sample to days in which the threshold is within reach, given the team's inherited output. Next, in Figures 5(b), (e) and (h), we plot the distribution of each day's *actual* output relative to the three thresholds (T, 1.3T, and 1.6T respectively). We show these distributions separately for days that are the last of a calendar month versus all the other days on which T is within reach.[31] Sales are measured in $100 bins relative to T, so a team that beat the target by $350, for example, would fall in the middle of bin 3. A team that made or beat the target by less than $100 would fall into bin 0.

Again, and consistent with the standard model, Figure 5(b) shows a pronounced difference between the output distributions on the last day of the month –which is the team's last chance to reach T– compared to all other days on which T is within reach: on the last day, a disproportionate share of teams sell just enough to beat the target by less than $100. Contrary to the standard model, however, Figures 5(e) and (h) also show clear bunching at the two kinks, 1.3T, and 1.6T, but *only* on last days of the month. Finally, the remaining parts of Figure 5 repeat this exercise for placebo kinks at 0.9T, 1.1T, 1.2T, 1.4T, 1.5T, and 1.7T. Here, the output distributions for last days versus all other days essentially coincide, with the possible exception of 1.5T.[32] Together, these results suggest that teams who find themselves within reach of a salient threshold on the last day of the month tailor their efforts to just reach that threshold.

4.3 Expanding the Control Group and Identification Strategy

The patterns in Figure 5 –which use non-last days on which a threshold is within reach as controls for last days on which a threshold is within reach– are suggestive, but they have limited statistical power due to the small number of non-last days on which a threshold is within reach. To raise our statistical power, this Section presents evidence suggesting that we can use *all* the other days in a store (all non-last days plus last days with no threshold within reach) as controls in our econometric analysis. To that end, Figure 6 compares the daily output distributions on last days when *any* threshold is within reach (our treatment days) to two other candidate control groups: last days on which no threshold is within reach, and non-last days. Figure 6 shows that the distributions on these two candidate control groups coincide very closely: a Kolmogorov-Smirnov (KS) test for no difference has a p-value of 0.364. In contrast, the gray dotted curve in Figure 6 (from last days *with* a threshold within reach) drifts away from the previous two curves. In particular, it is less likely to have extremely low sales of around [$0, $200), and more likely to exhibit sales above $800. Indeed, even when we restrict attention to last days, the sales distribution when a threshold is within reach is different from the distribution when the threshold is not reachable (a KS test for

[31] Note that the daily sales distributions in Figure 5 tend to be centered a little below T. This a consequence of our three-day window for a threshold being within reach: With such a window, the threshold is on average about 1.5 days of sales away on the morning of the current day.

[32] The graph for 1.5T suggests that a few teams may have aggressively pursued the highest kink at 1.6T and fallen just short. We caution, however, that sample sizes for (actual or placebo) kinks at 1.5T and higher are limited due to the small number of team-days on which these high output levels are within reach.

no difference has a p-value of 0.013).

Motivated by Figures 5 and 6, our main econometric approach will use all non-last days of the month, plus last days on which no threshold is within reach, as a control group for treated days. Intuitively, the distinction between treated and non-treated days is whether there is a clear connection between a sales team's effort on that day and the team's chances of attaining a potentially meaningful monthly sales threshold.[33] A key advantage of having a large group of control days is that –in contrast to some earlier bunching estimators– we can estimate the effect of the presence *and* location of a threshold on the entire daily sales distribution, without needing to assume any kind of smoothness. Instead, this distribution is represented by a finite number of sales bins, with an arbitrary baseline density attached to each of them. In this sense, our estimation approach is similar to Abeler et al. (2011), who study how the excess density in workers' performance levels around an induced reference point *moves* when that reference point is manipulated by the experimenters.[34]

A natural question concerning our approach is whether the variation we exploit –the precise location of the output bin in which a threshold falls on the last day of a month– is 'as good as random'. We offer three observations on this question. First, we note that the usual sorts of omitted variables that would affect an analysis of daily mean sales –for example a persistent shock to store output (perhaps due to a sales promotion) that could both bring a threshold within reach and raise the store's chances of exceeding it– do not apply to our approach. Such shocks can raise mean sales on adjoining days, but they have no obvious effect on the probability that both the threshold and actual sales fall into the same \$100 sales *bin* on the last day of the month. A second and perhaps more plausible source of endogeneity actually works in our favor. Suppose that –despite the unpredictable nature of daily sales– some teams successfully adopted a forward-looking strategy of tailoring their sales on the last few days of the month so as to put a particular threshold within 'reasonable' reach on the final day.[35] If this type of forward-looking behavior accounts for our main results, it only strengthens our primary hypothesis, which is that teams tailor their sales efforts to 'just reach' salient thresholds in the monthly pay schedule. Third, in addition to the evidence presented in this Section, we conduct a number of specification tests to show that our results are not sensitive to the use of a number of alternative control groups.

[33] As already noted, such a connection is absent on non-treated days for a variety of reasons, including the highly unpredictable nature of daily store sales documented in Table 1: early in the month, the connection between effort exerted on a particular day and the chances of reaching any particular monthly threshold is very weak. Other reasons why such a connection could be absent on a control day are that (a) the threshold has already been attained, (b) the threshold is out of reach, or (c) today is not the team's last chance to attain the threshold for the month.

[34] Our approach is also similar is spirit to Benson (2015), who creates a 'treatment bubble' of managers whose cumulative performance near the end of a pay period makes their current actions decisive for attaining a bonus.

[35] More specifically, suppose that teams try to get the threshold into a region we call the 'sweet spot' on the last day. A threshold is in the sweet spot if it is attainable with between one and two typical days' worth of sales; in Section 6.3 we find that thesholds in this location account for much of the bunching we observe.

5 Empirical Specification

As noted, our empirical strategy compares the distribution of a store's sales on last days of the month when a threshold is within reach to the same store's sales distribution on all other days. To simplify the discussion, we will say that a threshold is *attainable* on a given day if the threshold is within reach *and* the day is the last of a calendar month. Notably, in addition to the presence of an attainable threshold, we use the variation in the *position* of these thresholds that is generated by quasi-random variation in the store's inherited output on the morning of the last day of the month to identify the thresholds' effects. To implement our approach, we create a set of $100 daily sales bins for every store, which run from zero to the maximum daily sales we see for that store in our data period.[36] With 103 stores in our sample, this yields 34,863 daily store sales observations and 2,778,836 store$\times$day$\times$bin observations.

5.1 Estimation without Covariates

To illustrate how our approach works, we first consider an example where the reward schedule has a single threshold (for example, at the sales target, T), and no adjustment is made for covariates– for example, factors like weekends versus week days that affect the distribution of sales. In this case we would estimate the following linear probability model:

$$S_{imtb} = \alpha + \beta \cdot K^T_{imt,0} + \sum_{j=-5}^{-1} \gamma_j \cdot K^T_{imt,j} + \sum_{j=1}^{5} \gamma_j \cdot K^T_{imt,j} + I_i + \epsilon_{imtb}. \quad (1)$$

Our dependent variable, S_{imtb} takes a value of one if store i's actual sales on day t fall into bin b, and zero otherwise. Our main regressor of interest, $K^T_{imt,0}$ is a dummy variable that equals one if T is *attainable* on that day *and* the current sales bin contains the target T. In addition, we include indicators for the five sales bins that are just above or below the target's bin. Thus, the coefficient β estimates the additional likelihood that a store's sales output falls into bin b when that bin contains an attainable target. The ten coefficients γ_{-5} through γ_5 identify any excess or missing mass in the output bins surrounding an attainable target. Store fixed effects I_i are included to account for the fact that the number of output bins varies across stores. Finally, to measure the effects of the two 'pure' kinks (1.3T and 1.6T) affect the density of sales, we re-estimate equation (1) for the analogously defined regressors, $K^{1.3T}_{imt,0}$ and $K^{1.6T}_{imt,0}$. Estimates of equation (1) for T, 1.3T and 1.6T are presented in columns 1-3 of Table 3.

[36] Since stores differ in size and performance, this yields a different number of bins for each store. While this might seem cumbersome, forcing every sales bin to be $100 wide has the advantage that we can unambiguously interpret 'hitting' the bin that meets the team's monthly target as being within $100 of the target. Defining bins based on absolute store sales (rather than relative to a target) also respects the fact that control variables, such as day of the week and holidays, operate on the absolute sales distribution. This approach also accommodates noise in the production process: By classifying any level of sales in the same $100 bin as being 'at' the target, it includes teams that 'tried but just missed' as 'bunchers'.

5.2 Adding Covariates

One limitation of equation (1) is that it does not account for factors like holidays and the day of the week to affect the sales distribution. While we cannot think of an obvious reason why these factors will be correlated with the presence of an attainable sales target *in a particular bin* on the final day of a month, these factors are highly correlated with total daily sales, so it seems reasonable to control for them. Another limitation of equation (1) is the fact that it does not contain *bin* fixed effects. Thus it implicitly assumes that the baseline sales distribution (i.e. the distrbution on days when no threshold is attainable) is uniform.

To address the above issues, we adopt a set of fixed effects that allow each store to have its own baseline sales distribution with unrestricted shape. In addition we allow each store's baseline distribution to have a different shape on each day of the week. Both of these goals are accomplished by giving each store its own set of DOW×bin dummies. Similarly, to account for the substantial improvement in sales during holidays, we control for Holiday×bin fixed effects, where Holiday is defined as a categorical variable that identifies 10 major holidays when people shop heavily in China.[37] This yields the following regression:[38]

$$S_{imtb} = \alpha + \beta \cdot K_{imt,0} + \sum_{j=-5}^{-1} \gamma_j \cdot K_{imt,j} + \sum_{j=1}^{5} \gamma_j \cdot K_{it,j} + \Phi_{imt_{DOW}b} + \Theta_{t_H b} + \epsilon_{imtb}, \qquad (2)$$

where t_{DOW} indexes the days of the week and t_H indexes ten major holidays. All told, the control variables $\Phi_{it_{DOW}b}$ and $\Theta_{t_H b}$ comprise 56,252 store×DOW×bin dummies and 3,971 holiday×bin dummies. Standard errors are clustered by store.[39] With heterogeneity across time and stores being accounted for in this way, equation (2) again tests whether sales are more likely to fall into a bin that exactly attains a threshold, compared to what we would expect that store×day to produce otherwise. Estimates of equation (2) for T, 1.3T and 1.6T are presented in columns 4-6 of Table 3.

5.3 Interpretation and Magnitudes

The β coefficients in equations (1) and (2) give us the excess density in sales bins that contain an attainable threshold. In order to interpret those magnitudes it is helpful to have an estimate of the baseline density that we would expect to see in the bins that 'tend to' contain a target or kink, when no target or kink is present. In Appendix B, we describe in detail how these densities are created. Briefly, for each of T, 1.3T, and 1.6T, we calculate for each store the probability the threshold falls

[37] The holidays include New Year, Chinese New Year, May Day, National Day, Father's Day, November 11th, December 12th, and three other traditional Chinese holidays.

[38] Like equation 1, equation 2 focuses on the monthly target, T. Estimating equations for the two kinks, 1.3T and 1.6T, are defined analagously.

[39] Clustering by stores accommodates a potential concern with estimating a complete distribution of sales for every store: Because sales must fall into exactly one of a store's (say) 40 bins on every day, there will a mechanical correlation of errors across the bins within every store×day cell. Thus it seems essential to cluster at least at the store×day level. Clustering at the store level should handle this issue, plus more general concerns such as autocorrelation of errors within stores over time.

into every one of the store's output bins. We then use these probabilities as weights to calculate the expected density in the output bins where an attainable threshold would 'typically' land on control days. These densities are 10.3, 7.7 and 5.9 percentage points for T, $1.3T$, and $1.6T$ respectively. (Because the higher thresholds are attained less frequently, their baseline densities are lower.) We will use these three baseline levels to interpret magnitudes in the following Section.

6 Results

6.1 Main Results

Estimates of equations (1) and (2) are presented in Table 3. Notably, across all six columns, we find substantial bunching at the target T, and also at the two pure kinks $1.3T$ and $1.6T$. In the absence of controls, column (1) shows that a \$100 sales bin that exactly contains an attainable target T is 17.1 percentage points more likely to be exactly achieved than the same bin when it does not contain an attainable target. In the presence of controls (column 4), this estimated coefficient is 9.19 percentage points. Compared to the baseline density in the sales bins that typically contain T (10.3 percentage points) this estimate corresponds to an 84 percent increase. Column (4) also shows statistically significant missing mass in the bin just below target, equal to 4.7 percentage points, or 46 percent. Columns 5 and 6 show estimates from the same regression for coefficients corresponding to the two kinks, $1.3T$ and $1.6T$. These show excess density of 11.4 and 14.9 percentage points respectively, which represent increases of 149 and 252 percent. Overall, in all three columns (4) through (6), we find sizable and significant bunching effects at targets and kinks, but –with the exception of the bin just below the target– no strong evidence of nearby missing masses.

To shed additional light on where the additional density at attainable thresholds in Table 3 is coming from, Table 4 estimates the following specification:

$$S_{itb} = \alpha + \beta \cdot K_{it,0} + \Gamma_{\ominus} \cdot K_{it,\ominus} + \Gamma_{\oplus} \cdot K_{it,\oplus} + \Phi_{it_{DOW}b} + \Theta_{t_H b} + \epsilon_{itb}, \tag{3}$$

Specifically, we create a dummy variable, $K_{it,\ominus}$, equal to 1 for *all* of a store's bins strictly below a bin that contains an attainable threshold (on the day that threshold is attainable), and a dummy variable, $K_{it,\oplus}$, equal to 1 for all the bins strictly above it. These dummies give the effect of output falling in *any* of the included \$100 bins when that kink is attainable. We then scale up these two estimates by the number of \$100 bins they include to get the estimated change in total density below and above the threshold. Together with our estimate for the threshold itself, these three estimates sum to one, and summarize *how the entire sales distribution shifts* on days when an attainable threshold is present. Consistent with the canonical model, the presence of an attainable target T in a bin shifts 17.1 percentage points of density out of the region below the target. Of this, 9.1 percentage points is moved into the bin containing the target, and the remaining 8.3 percentage points is moved into output levels strictly above the target. Similarly, the presence of an attainable pure kink ($1.3T$ or $1.6T$) shifts 18.5 (21.6) percentage points of density out of the region below the kink. 11.4 (14.9) of this is shifted into bin containing the kink, with the remainder moving to strictly higher levels.[40] This confirms that –at least on days where they are both within reach and at risk– both bonuses *and* concave kinks in the pay schedule appear to be highly motivating for workers, even in a team context where one might expect free-riding problems to arise.

Finally, to quantify the change in mean sales on days when a threshold is attainable, we simply regress total daily sales on a dummy for whether the team's target (T) or one of its two kinks ($1.3T$ or $1.6T$) were relevant on that day.[41] As shown in Table 5, sales output is $240 (41 percent) higher on days where a target is attainable, and $191 (33 percent) higher when $1.3T$ is attainable, than on comparable days with no attainable threshold. For the next kink at $1.6T$, the point estimate is $258, but is statistically insignificant.

6.2 Robustness Checks

In our main analysis, we treated a threshold as attainable if it is within no more than three times the store's mean daily sales during the previous days of the current month. In Appendix Table C1, we revise this definition and define a threshold to be attainable if it is within three times the store's mean daily sales during the previous same *days of the week* in the current month. Compared with the original definition, this definition includes some thresholds from high-sales days (like weekends) that were previously considered unattainable, and excludes some observations from weekdays since the attainable range is now defined more narrowly for these days. As shown in Table C1, the bunching effects for T, $1.3T$, and $1.6T$ are robust to this definition.

In our main analysis we maximized our sample by using all the non-last days in a month as controls for last days when a target or a kink is attainable. However, as noted earlier it is possible that the behavior of forward-looking teams in the days leading up to the end of the month is strategic. While this is not a source of concern for our main results –such strategic behavior would only provide additional support for our main hypothesis that teams tailor their effort levels to 'just' attain meaningful monthly sales thresholds– it is nevertheless of interest to assess whether this more forward-looking type of behavior might be affecting our results. To address this issue, Appendix Table C2 replicates Table 3, dropping the seven days preceding the final day of the month from the control group. The estimates are very similar, suggesting that team behavior in the week leading up to the final day is not qualitatively different from earlier days in the month. A second possible concern with our control group is that –despite the evidence in Figure 6 which suggests

[40] These increases in density strictly above the kinks are not significantly different from zero, however.

[41] As noted earlier, we are less confident in these daily mean sales regressions than in our main estimates. This is because unmeasured shocks to team sales which last more than one day (such as unobserved sales and promotions) could cause the team's target to be reached, creating bias from reverse causation. That said, Table 5 mitigates this concern by including controls for the store's sales in each of the three days preceding the end of the month.

otherwise– the last days of a month are so unique that no other days from the same month are good comparators. To address this possibility, Table C3 replicates Table 3 using *only* last days when no threshold is attainable as controls for last days when such a threshold is attainable.[42] Once again, the results are very similar to Table 3.

6.3 Heterogeneity Analysis

Thoughout our analysis so far, we have treated all thresholds within three days' worth of sales as within reach. However, if our sales teams truly treat the thresholds as goals, it is possible that the thresholds with the largest effects on sales are the ones that are neither too easy nor too hard to attain. To see if that is the case, Table 6 disaggregates Table 3's thresholds into three types (using the full specification in columns 4-6):

- 'easy' goals (when the threshold is within 1 day of mean store sales),
- 'sweet spot' goals (when the threshold is between 1 and 2 days of sales), and
- 'stretch' goals (when the threshold is more than 2 days of sales away).

If targets and kinks function as implicit goals, then when a goal is very close, we would expect less bunching, since most teams would go beyond it even with 'normal' effort. Similarly, when a goal is very far away, we would also expect less bunching, since the goal may be impossible to reach for most teams. But if the goal is right in the 'sweet spot' (challenging but still attainable), it could be especially effective in motivating agents to reach it. Consistent with these expectations, Table 6 shows that bunching is strongest when a sales target *or* commission kink falls into the 'sweet spot' that is neither too easy nor too hard to attain.

Appendix D shows the results of three additional heterogeneity analyses, beginning with the effects of recent turnover events at the store. Recent turnover may affect the amount of bunching if less-intact teams find it harder to co-ordinate on a shared sales goal; this is especially relevant to Firm A because most departures occur near the end of the month (Kuhn and Yu, 2021). However, as shown in columns (1)-(2) of Table D1, the likelihood of bunching at both the target and kinks is not significantly different in months when a turnover has occurred, compared to other months. In columns (3)-(4), we repeat this analysis for turnover in the *preceding* month, and we still find no significant difference in the amount of bunching at any of the thresholds.

Next, we examine whether the amount of bunching varies with the size of the team. While we have modeled the behavior of Firm A's sales teams as though they were individual workers, it is interesting to ask how issues like free riding and co-ordination problems might affect output bunching patterns. On the one hand, if larger teams find it hard to co-ordinate their activities to

[42] To economize on degrees of freedom, Table C3 replaces the Bin×Store×DOW fixed effects in Table 3 by Bin×Store FEs plus Bin×Holiday fixed effects. (The Table 3 specification has essentially only 12 degrees of freedom and exhibits very high stardard errors).

obtain a threshold public good like the bonus at the target, we should see less bunching at T among larger teams. On the other hand –and closely related to our symbolic rewards model– teams might take advantage of salient points in the sales schedule as focal points to solve co-ordination and free rider problems. In this case, attaining a clearly defined, shared goal could generate utility for team members and create *more* bunching at all three thresholds in larger teams. As shown in Table D2, bunching at both the target and the kinks is stronger in larger teams; these differences are large in magnitude and highly statistically significant. We conclude that all three thresholds might, indeed, be functioning as a coordination devices for team production (Heath et al., 1999; Allen et al., 2017).

Finally, we ask whether experienced sales teams are more or less likely to bunch at kinks than other teams. If bunching is a behavioral bias that reduces experienced utility (as in our loss-aversion model), experienced teams might learn *not* to bunch; this is especially the case for the two kinks because classical utility-maximizing behavior involves *avoiding* kinks. In contrast, if –as suggested above– bunching at kinks is an effective way for teams to motivate their members, experienced teams might bunch more than others. As shown in Table D3, the amount of bunching at both the target and the kinks increases with average firm tenure of the sales team; in both cases the interaction terms are large and statistically significant. Once again, the evidence suggests that teams may use the kinks as goals to co-ordinate their efforts.[43]

7 Conclusion

This paper has studied the performance of 103 small retail sales teams that face a nonlinear commission schedule that includes a target (with a bonus attached to it) plus two 'pure' concave kinks. We have shown that the presence of an attainable target *or* kink on a workday shifts the distribution of daily sales from below the target or kink towards higher levels. Thus, accelerating commission rates appear to be effective motivators, relative to a linear commission scheme at the pre-kink commission rate.[44]

A surprising finding of our paper is the bunching of performance around the two pure concave kinks. In the paper we show that a symbolic-rewards model accounts for this result better than a loss-aversion model. Notably, this symbolic aspect of the pure, convex kinks in Firm A's pay schedule appears to be strong enough to override the purely economic incentives that motivate teams to avoid, rather than congregate at those points. We also find that bunching at all three of the thresholds in teams' pay schedules is stronger in larger, experienced teams than smaller,

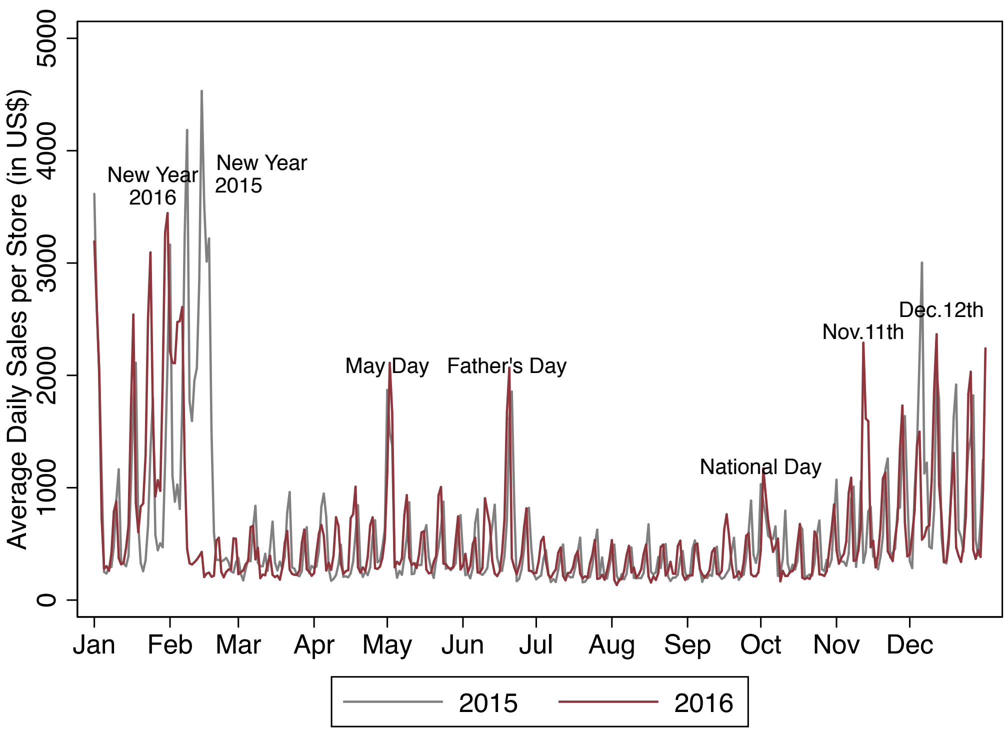
[43]Motivated by these findings, we were able to converse with a small number of team members, who confirmed that some team leaders promote the thresholds as informal team goals for the month.

[44]While we cannot formally show that adding such accelerators raises Firm A's profits, we note three factors that make this more likely. First, while adding accelerating commissions will raise compensation costs for workers who qualify for them, they are much less costly than raising the value of a linear commission rate because they only apply to sales above the kink. Second, accelerating commission rates may also have attractive selection properties for firms: they provide a simple, rules-based way to raise the pay of high-ability workers without paying other workers more. Third, at least in our context, most of the negative side-effects of nonlinear incentives, such as timing gaming and price distortion, do not apply.

inexperienced ones. This suggests that the symbolic rewards associated with attaining different ranges of the output distribution may be best interpreted at shared *goals* that are used by teams to motivate and co-ordinate their members.

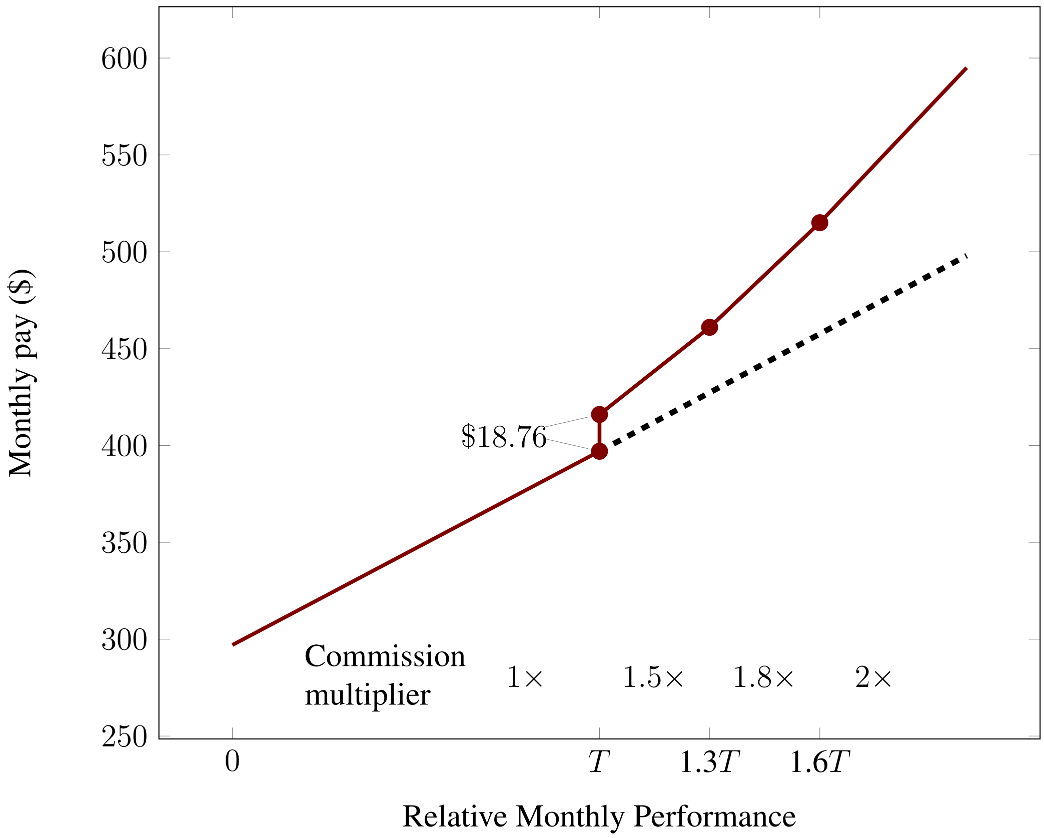
Finally, we note that Firm A's nonlinear pay scheme functions as part of system of personnel policies (Boning et al., 2007) that allow Firm A to operate in the very particular environment it faces. In this environment, all the sales employees are dispersed across 103 remote locations (stores), making it costly for Firm A to manage and monitor employees' daily activities. Management does, however, observe each store's total daily sales. Given these monitoring and informational constraints, Firm A has adopted a flexible, team-based policy that relies on peer monitoring among team members and internal coordination to improve team production. Team-based commissions not only incentivize teams to sell more; they also reward team members for monitoring each other, for training and co-operating with each other, for selecting able new employees, and for encouraging under-performing employees to leave (Kuhn and Yu, 2021). The *target* in this commission scheme serves as an effective way to communicate management's month-and-store-specific expected performance levels to the teams. The concave kinks, or *accelerators*, provide additional financial incentives at relatively low costs to the firm and appear to function as symbolic rewards as well. We hope that these findings will encourage further research on team-based accelerators, and on symbolic rewards for *absolute, team-based* performance as effective motivators that lack the negative spillover effects of other reward schemes.

Figure 1: Average Daily Sales per Store by Calendar Day



Notes: This figure plots average daily sales per store on every calendar day in 2015 and 2016. The labelled spikes correspond to holidays or major shopping events. New Year denotes the Chinese Lunar New Year. November 11th and December 12th are the major shopping events in China, similar to the Black Friday or the Cyber Monday.

Figure 2: Pay Scheme



Notes: This figure plots Firm A's pay scheme in 2016, based on the average amount of base payment and commission rate. Relative monthly performance is defined as the monthly output in 2016 relative to the monthly target, i.e. the same store's monthly output during the same calendar month in 2015. Employees' monthly pay includes a base payment and a commission payment. Commission payment is based on the store's total monthly output, multiplied by a pre-determined baseline commission rate ($1\times$). If the store has met its monthly goal, then each team member is rewarded with a lump sum bonus of $18.76 (100 CNY). On sales above T, commission rate is 50% higher than the baseline rate ($1.5\times$). There are another two *pure* kinks in the pay schdule: on sales between $1.3T$ and $1.6T$, the commission rate is 80% higher than the baseline rate ($1.8\times$); on sales exceeding $1.6T$, the commission rate is 100% higher than the baseline rate ($2\times$). For reference, the dashed line shows the linear schedule under the baseline commission rate.

Figure 3: A Template of Pay Slip

工资条															
									年 月 日			编号			
序号	姓名	应发									应扣				实发数
		基本工资	绩效工资				绩效奖金	电话补贴	加班工资	小计	绩效扣款	缺勤天数	代扣	小计	
			标准	优良	优秀	杰出									
xx	XXX	2000	1000	450	18		100			3568			-200	-200	3368

月度绩效： 优秀

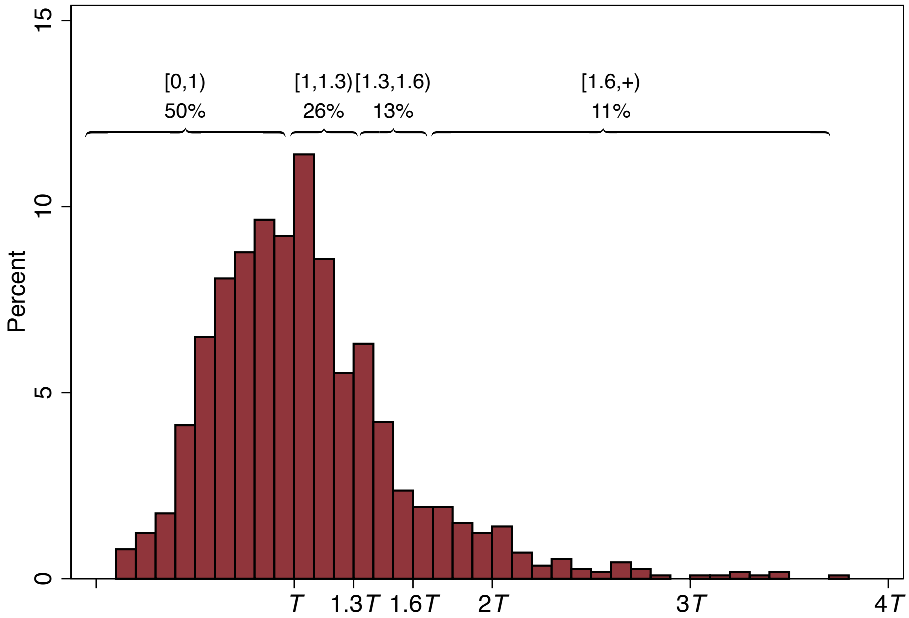
Pay Slip															
									Date			No.			
#	Name	Earnings									Deductions				Net Payment
		Base	Commission				Bonus	Cell Phone Allowance	Overtime Pay	Total	Penalty	Absenteeism	Other Deductions	Total	
			Standard	Good	Excellent	Superb									
xx	XXX	2000	1000	450	18		100			3568			-200	-200	3368

Monthly Asseessment: Excellent

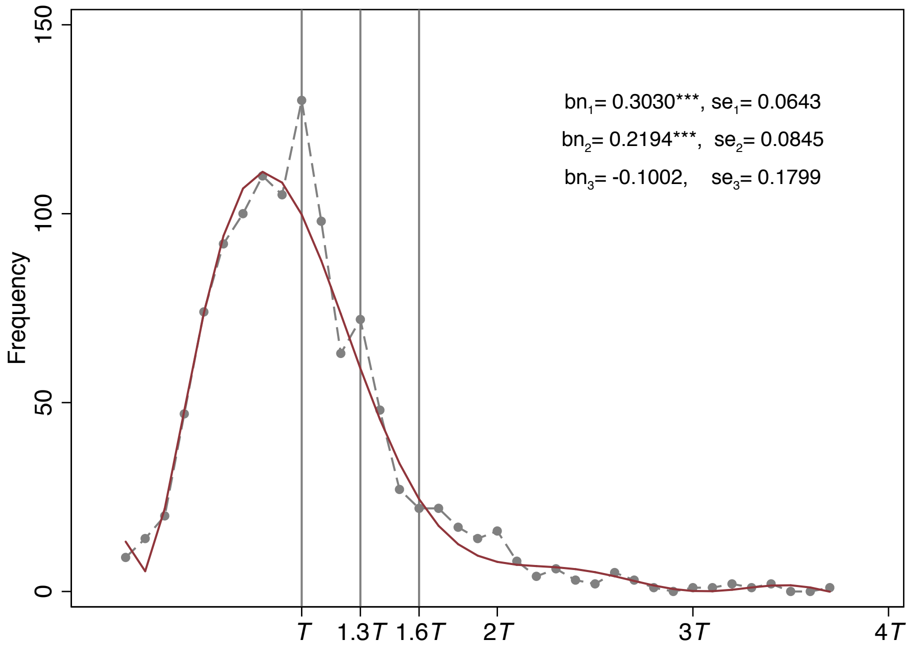
Notes: This figure presents a template of the pay slip to salespeople. The commissions are presented separately with the applicable commission rates, and the ranges are titled as “Standard”, “Good”, “Excellent”, and “Superb”. On the bottom of the pay slip, the monthly assessment takes the highest nonempty title.

Figure 4: Relative Monthly Performance

(a)

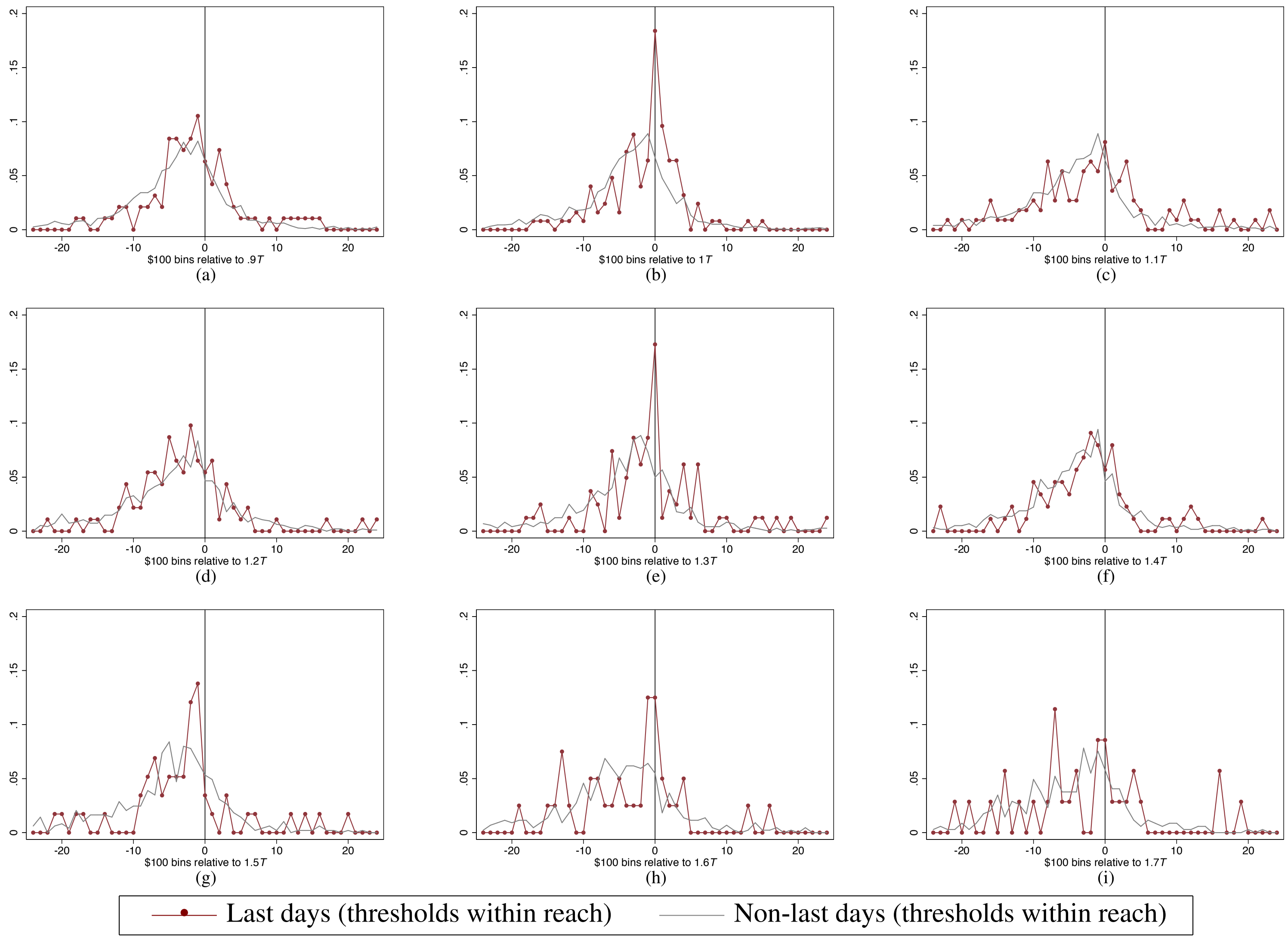


(b)



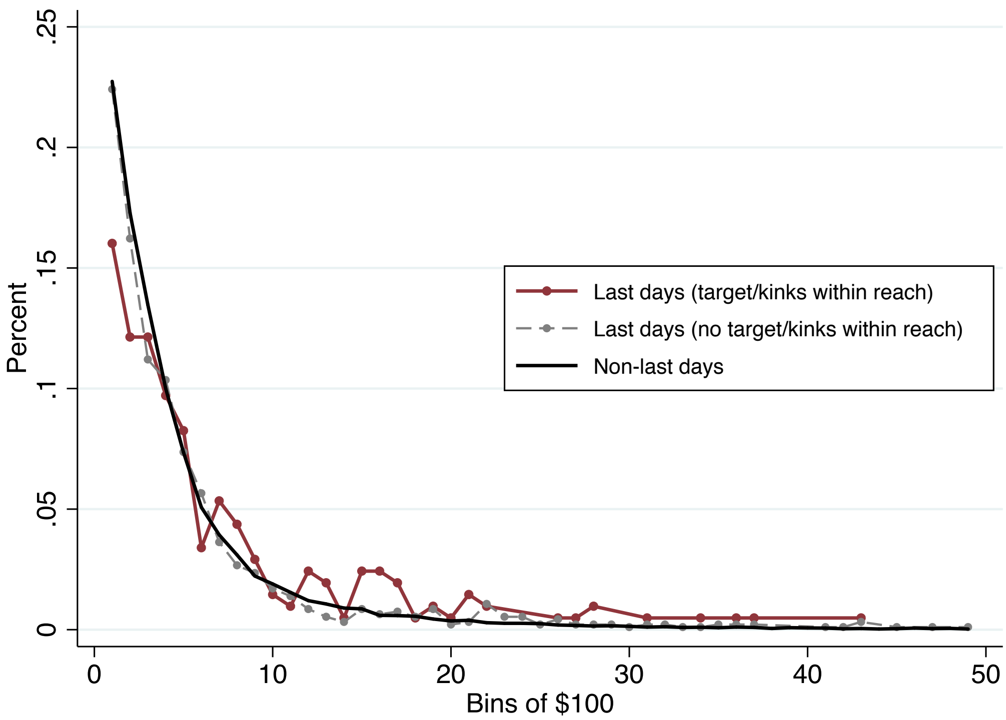
Notes: Figure 4(a) plots the histogram of relative monthly performance, i.e. the monthly output in 2016 relative to the same store's output in 2015 during the same calendar month. Figure 4(b) plots the actual number of observation in each bin (connected dotted line), and the estimated number of observations in each bin (dashed curve) from a high-degree polynomial.

Figure 5: Excess (or Deficient) Sales Relative to the threshold



Notes: These figures plot the histogram of excess (or deficient) sales relative to the corresponding kinks, measured in bins of $100, separately for last days and non-last days. For better visual display, only the range [-$2,500, $2,500] is presented. Figures (b), (e), and (h) in the center column are plotted for kinks T, $1.3T$, and $1.6T$, respectively; figures on the left and right columns are plotted for placebo kinks.

Figure 6: Distribution of Daily Sales Output



Notes: This figure plots the distribution of daily sales output in bins of $100. For better visual display, only the range [$0, $5,000] is presented. The distribution is calculated separately for each of the following three samples: last days with kinks relevant (dotted solid curve), last days with no kinks relevant (diamond-marked dashed curve), and non-last days (solid curve).

Table 1: Descriptive Statistics

	Mean	SD	Median	N
Panel A: Product price (in US $)	51.72	49.79	30.94	437
Accessories	16.41	9.77	20.03	20
Shirts and Polos	27.66	10.05	27.81	142
Pants	27.81	10.84	24.69	129
Sweaters	32.66	16.06	28.59	19
Jackets	101.09	34.23	82.50	91
Suits	180.28	53.94	180.94	36
Panel B: Daily sales (in US $)	582	1241	273	34,863
Target size=2	274	368	172	6,376
Target size=3	634	1446	266	18,728
Target size=4	680	1150	382	9,759
Panel C: Relative Monthly Performance	1.07	0.56	1.00	1,143
Target size=2	0.97	0.45	0.91	209
Target size=3	1.10	0.62	1.01	614
Target size=4	1.07	0.49	1.00	320
Panel D: Monthly compensation (in US $)	545	156	512	4,176
Target size=2	507	96	488	446
Target size=3	549	171	513	2,126
Target size≥4	550	145	525	1,604

Notes: Product prices presented in Panel (A) are from a sample of items sold in September, 2016; prices reflect original tag prices and discounted prices. Target size in Panels (B) and (C) is observed from the annual sales plan, at store-year level. For newly-opened stores whose target size is not available in the current year, we use team size 30 days after the opening instead. Monthly compensation includes a base salary and a commission component based on team performance, along with the social security payments. Monthly compensations are missing for 6% of employee-month observations. Tenure experience in Panel (D) measures employees' firm tenure in years.

Table 2: Model Predictions

	Change in Mass		
	Below the Threshold	**At the Threshold**	**Above the Threshold**
A: Predictions at the 'pure' kinks (1.3*T* and 1.6*T*)			
1. Standard Model	–	–	+
2. Symbolic Rewards (SR) Model	–	+	+
3. Reference Point (RP) Model (weak loss aversion)	–	–	+
4. Reference Point (RP) Model (strong loss aversion)	–	+	–
B: Predictions at the target (*T*)			
1. Standard model	–	+	+
2. Symbolic Rewards (SR) Model	–	+	+
3. Reference Point (RP) Model (weak loss aversion)	–	+	+
4. Reference Point (RP) Model (strong loss aversion)	–	+	–

Notes: 1. Changes in mass refer to the difference in total density between two situations: (a) when a target or kink is present, versus (b) when the worker faces a linear commission at the (lower) rate that prevails to the left of the threshold.
2. "At the threshold" refers to the bin containing the threshold (and to immediately neighboring bins if teams can't perfectly target). "Below" (above) the threshold refer to the regions of output space outside this interval. Thus, for example, (–,+,+) bunching at the threshold (and possibly a neighboring bins), missing mass below this bunching region, and excess mass above it.
3. The shaded rows indicate that the prediction is consistent with our empirical estimates.
4. Predictions are derived in Appendix A.

Table 3: Estimates of Excess Density in and around Attainable Output Thresholds

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month					
	Specification 1: Raw Estimates			Specification 2: FEs Controlled		
	T	1.3T	1.6T	T	1.3T	1.6T
$K_{imt,-5}$	0.0481	0.0340	-0.0125***	-0.0211	-0.0362	-0.0536***
	(0.0298)	(0.0326)	(0.0012)	(0.0261)	(0.0337)	(0.0125)
$K_{imt,-4}$	0.0375	0.0463	0.0986	-0.0355	-0.0285	0.0672
	(0.0240)	(0.0328)	(0.0597)	(0.0239)	(0.0250)	(0.0589)
$K_{imt,-3}$	0.1303***	0.0401	0.0232	0.0396	-0.0331	0.0010
	(0.0396)	(0.0295)	(0.0341)	(0.0380)	(0.0324)	(0.0264)
$K_{imt,-2}$	0.0568**	0.0935**	0.0178	-0.0290	0.0077	-0.0522
	(0.0285)	(0.0424)	(0.0301)	(0.0262)	(0.0382)	(0.0402)
$K_{imt,-1}$	0.0321	0.0675**	0.0415	-0.0469**	0.0096	-0.0229
	(0.0329)	(0.0424)	(0.0740)	(0.0184)	(0.0291)	(0.0274)
$\boldsymbol{K_{imt,0}}$	**0.1707*****	**0.1923*****	**0.1980*****	**0.0919*****	**0.1143****	**0.1494****
	(0.0329)	**(0.0424)**	**(0.0740)**	**(0.0328)**	**(0.0439)**	**(0.0665)**
$K_{imt,1}$	0.0944***	0.0357	0.0138	0.0394	-0.0069	-0.0202
	(0.0259)	(0.0237)	(0.0255)	(0.0270)	(0.0228)	(0.0255)
$K_{imt,2}$	0.0638**	0.0116	0.0401	0.0274	-0.0239	0.0134
	(0.0262)	(0.0167)	(0.0370)	(0.0250)	(0.0160)	(0.0381)
$K_{imt,3}$	0.0333*	0.0116	0.0138	0.0067	-0.0116	0.0008
	(0.0178)	(0.0167)	(0.0254)	(0.0158)	(0.0164)	(0.0235)
$K_{imt,4}$	0.0486**	0.0357	0.0401	0.0310	0.0208	0.0352
	(0.0208)	(0.0228)	(0.0349)	(0.0203)	(0.0228)	(0.0349)
$K_{imt,5}$	0.0028	0.0116	-0.0125***	-0.0023	0.0051	-0.0187***
	(0.0109)	(0.0165)	(0.0012)	(0.0107)	(0.0172)	(0.0048)
N	2697816	2693899	2691619	2697816	2693899	2691619

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Columns (1)-(3) report estimates of equation (1), which controls for store fixed effects. Columns (4)-(6) report estimates of equation (2), which controls for Bin×Store×DOW and Bin×Holiday fixed effects. Standard errors reported in parentheses are clustered at the store level. Each column represents a different regression, and the sample in each column comprises the control sample plus the treated sample for the threshold in question (T, 1.3T, or 1.6T). An observation is a bin x store x day cell. $\boldsymbol{K_{imt,0}}$ is an indicator variable, identifying the bin that exactly contains the threshold; the remaining 'K' variables indicate the five bins immediately below and above the threshold.

Table 4: Estimates of Excess Density below, at, and above Attainable Thresholds

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	Target (T)	First Kink ($1.3T$)	Second Kink ($1.6T$)
$K_{imt,\ominus}$	-0.1714***	-0.1851***	-0.2156***
	(0.0344)	(0.0427)	(0.0681)
$K_{imt,0}$	0.0919***	0.1143**	0.1494**
	(0.0328)	(0.0439)	(0.0665)
$K_{imt,\oplus}$	0.0828*	0.0718	0.0785
	(0.0414)	(0.0457)	(0.0698)
Bin×Store×DOW	Yes	Yes	Yes
Bin×Holiday	Yes	Yes	Yes
N	2697816	2693899	2691619

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Notes: This Table reports estimates of equation (2), which controls for Bin×Store×DOW and Bin×Holiday fixed effects. Standard errors reported in parentheses are clustered at the store level. Each column represents a different regression, and the sample in each column comprises the control sample plus the treated sample for the threshold in question (T, $1.3T$, or $1.6T$). An observation is a bin x store x day cell. $\boldsymbol{K_{imt,0}}$ is an indicator variable, identifying the bin that exactly contains the target or the kink. $K_{imt,\ominus}$ identifies *all* the bins below $\boldsymbol{K_{imt,0}}$, while $K_{imt,\oplus}$ identifies all the bins above it.

Table 5: Daily Mean Sales Regressions

	Dependent variable: Daily Sales (in US$)		
	T	$1.3T$	$1.6T$
Last day contains an attainable threshold	239.7**	191.3**	257.7
	(100.9)	(91.0)	(336.4)
Store×DOW	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
N	34863	34863	34863

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Notes: The dependent variable is the daily sales output; an observation is a store x day cell. The regressor of interest is an indicator variable, taking a value of 1 if the current day contains an attainable threshold of the type indicated (T, $1.3T$, or $1.6T$). All regressions control for the store's sales in each of the three days preceding the end of the month, Store×DOW fixed effects, and Holiday fixed effects. Standard errors reported in parentheses are clustered at the store level.

Table 6: Heterogeneity Examination – Easy Goals, 'Stretch' Goals, and Goals in the 'Sweet Spot'

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month								
	Easy [0, 1)			**Sweet [1, 2)**			**Stretch [2, 3)**		
	T	$1.3T$	$1.6T$	T	$1.3T$	$1.6T$	T	$1.3T$	$1.6T$
$K_{imt,-5}$	-0.0660***	-0.1516**	-0.1366***	-0.0404	-0.0224	-0.0509**	-0.0008	-0.0256	-0.0314**
	(0.0247)	(0.0699)	(0.0277)	(0.0356)	(0.0625)	(0.0222)	(0.0415)	(0.0435)	(0.0122)
$K_{imt,-4}$	-0.0651**	-0.0759**	-0.0999**	-0.0721**	-0.0249	0.0610	0.0013	-0.0202	0.1189
	(0.0258)	(0.0332)	(0.0445)	(0.0277)	(0.0569)	(0.0906)	(0.0436)	(0.0295)	(0.0962)
$K_{imt,-3}$	0.0049	-0.0878***	0.1425	0.0216	-0.0876***	-0.0318***	0.0663	0.0234	-0.0161*
	(0.0716)	(0.0271)	(0.1737)	(0.0647)	(0.0200)	(0.0101)	(0.0486)	(0.0650)	(0.0095)
$K_{imt,-2}$	-0.0819*	-0.1145***	-0.2128***	-0.0082	-0.0055	0.0224	-0.0253	0.0736	-0.0191***
	(0.0467)	(0.0258)	(0.0569)	(0.0552)	(0.0536)	(0.0976)	(0.0222)	(0.0585)	(0.0060)
$K_{imt,-1}$	-0.0741*	0.0031	0.0148	-0.0678**	0.0200	-0.0603**	-0.0066	0.0059	-0.0258**
	(0.0434)	(0.0682)	(0.0847)	(0.0275)	(0.0532)	(0.0270)	(0.0260)	(0.0336)	(0.0118)
$K_{imt,0}$	**0.0735**	**0.0939**	**0.1824**	**0.1188****	**0.1882****	**0.1410***	**0.0848***	**0.0740**	**0.1255**
	(0.0653)	**(0.0853)**	**(0.1127)**	**(0.0597)**	**(0.0843)**	**(0.0829)**	**(0.0447)**	**(0.0582)**	**(0.0920)**
$K_{imt,1}$	-0.0067	-0.0391	-0.0099	0.1403**	0.0076	-0.0372***	-0.0137***	0.0144	-0.0164*
	(0.0492)	(0.0465)	(0.0704)	(0.0611)	(0.0414)	(0.0136)	(0.0026)	(0.0345)	(0.0093)
$K_{imt,2}$	0.0455	-0.0364	0.0200	0.0189	-0.0331***	0.0504	0.0156	-0.0035**	-0.0217**
	(0.0611)	(0.0442)	(0.0764)	(0.0332)	(0.0066)	(0.0915)	(0.0239)	(0.0017)	(0.0106)
$K_{imt,3}$	0.0270	-0.0038	-0.0451***	0.0027	-0.0231***	0.0690	-0.0122***	-0.0100***	-0.0103*
	(0.0369)	(0.0444)	(0.0104)	(0.0235)	(0.0068)	(0.0806)	(0.0028)	(0.0030)	(0.0059)
$K_{imt,4}$	0.0231	0.0743	0.1167	0.0550	-0.0141***	-0.0154**	0.0151	-0.0059**	-0.0007
	(0.0370)	(0.0627)	(0.1023)	(0.0393)	(0.0044)	(0.0076)	(0.0237)	(0.0027)	(0.0014)
$K_{imt,5}$	-0.0122	-0.0046	-0.0337***	-0.0117***	0.0265	-0.0115**	0.0187	-0.0025	-0.0098
	(0.0228)	(0.0337)	(0.0105)	(0.0025)	(0.0405)	(0.0046)	(0.0232)	(0.0016)	(0.0063)
Bin×Store×DOW	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bin×Holiday	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2691267	2689892	2689126	2690917	2689934	2689018	2691086	2689527	2688929

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
This Table replicates columns (4)-(6) of Table 3 for three different types of thresholds– easy, 'sweet spot', and 'stretch'. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–92.

Adamson, B., Dixon, M., and Toman, N. (2014). Why individuals no longer rule on sales teams. *Harvard Business Review*.

Agarwal, S., Fang, X., Goette, L., Sing, T. F. S., Schoeb, S. S., Tiefenbeck, V., Staake, T. S., and Wang, D. (2017). The role of goals in motivating behavior: Evidence from a large-scale field experiment on resource conservation. *unpublished manuscript, University of Bonn*.

Allen, E. J., Dechow, P. M., Pope, D. G., and Wu, G. (2017). Reference-dependent preferences: Evidence from marathon runners. *Management Science*, 63(6):1657–1672.

Ashraf, N., Bandiera, O., and Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior and Organization*, 100:44–63.

Basu, A. K., Lal, R., Srinivasan, V., and Staelin, R. (1985). Salesforce compensation plans: An agency theoretic perspective. *Marketing Science*, 4(4):267–291.

Benson, A. (2015). Do agents game their agents' behavior? evidence from sales managers. *Journal of Labor Economics*, 33(4):863–890.

Berg, P., Appelbaum, E., Bailey, T., and Kalleberg, A. L. (1996). The performance effects of modular production in the apparel industry. *Industrial Relations: A Journal of Economy and Society*, 35(3):356–373.

Boning, B., Ichniowski, C., and Shaw, K. (2007). Opportunity counts: Teams and the effectiveness of production incentives. *Journal of Labor Economics*, 25(4):613–650.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics*, 134(3):1405–1454.

Charness, G., Kuhn, P., and Villeval, M. C. (2011). Competition and the ratchet effect. *Journal of Labor Economics*, 29(3):513–547.

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *The quarterly journal of economics*, 126(2):749–804.

Cohen, A. (2008). Package size and price discrimination in the paper towel market. *International Journal of Industrial Organization*, 26(2):502–516.

Dobronyi, C. R., Oreopoulos, P., and Petronijevic, U. (2019). Goal setting, academic reminders, and college success: A large-scale field experiment. *Journal of Research on Educational Effectiveness*, 12(1):38–66.

Douthit (1976). The use of sales quotas by industrial firms. *Journal of the Academy of Marketing Science*, 4:467–472.

Fortin, N., Lemieux, T., and Lloyd, N. (2021). Labor market institutions and the distribution of wages: The role of spillover effects. *NBER working paper*, (28375).

Freeman, R. B., Huang, W., and Li, T. (2019). Non-linear incentives and worker productivity and earnings: Evidence from a quasi-experiment. *National Bureau of Economic Research, working paper no 25507*.

Friebel, G., Heinz, M., Krüger, M., and Zubanov, N. (2017). Team incentives and performance: Evidence from a retail chain. *American Economic Review*, 107(8):2168–2203.

Gu, Z. and Yang, S. (2010). Quantity-discount-dependent consumer preferences and competitive nonlinear pricing. *Marketing Science*, 47(6):1100–1113.

Gubler, T., Larkin, I., and Pierce, L. (2016). Motivational spillovers from awards: Crowding out in a multitasking environment.

Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of political Economy*, 111(3):465–497.

Heath, C., Larrick, R. P., and Wu, G. (1999). Goals as reference points. *Cognitive psychology*, 38(1):79–109.

Holmstrom, B. and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, pages 303–328.

Joseph, K. and Kalwani, M. (1998). The role of bonus pay in salesforce compensation plans. *Industrial Marketing Management*, 27:147–159.

Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, 123(6):1227–1277.

Kleven, H. J. (2016). Bunching. *Annu. Rev. Econ*, 8:435–64.

Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *The Quarterly Journal of Economics*, 128(2):669–723.

Knez, M. and Simester, D. (2001). Firm-wide incentives and mutual monitoring at continental airlines. *Journal of Labor Economics*, 19(4):743–772.

Kosfeld, M. and Neckermann, S. (2011). Getting more work for nothing? symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3:1–16.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Kuhn, P. J. and Yu, L. (2021). How costly is turnover? evidence from retail. *Journal of Labor Economics*.

Lambrecht, A., Seim, K., and Skiera, B. (2007). Does uncertainty matter? consumer behavior under three part tariffs. *Journal of Marketing Research*, 26(5):698–710.

Larkin, I. (2014). The cost of high-powered incentives: Employee gaming in enterprise software sales. *Journal of Labor Economics*, 32(2):199–227.

Lawler, E. and Mohrman, S. A. (2003). Pay practices in fortune 1000 corporations. *WorldatWork Journal*, 12(4):45–54.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(5):1346–1361.

MyFirstSalesJob (2020). How to Structure Sales Commission Plans.

Ockenfels, A., Sliwka, D., and Werner, P. (2015). Bonus payments and reference point violations. *Management Science*, 61(7):1496–1513.

O'Connell, W. A. (1989). *Salesforce Compensation,Dartnell's 25th Survey*. Dartnell Corporation.

Oettinger, G. S. (2002). The effect of nonlinear incentives on performance: evidence from "econ 101". *Review of Economics and Statistics*, 84(3):509–517.

Owan, H., Tsuru, T., and Uehara, K. (2015). Incentives and gaming in a nonlinear compensation scheme. In *Evidence-based HRM: A Global Forum for Empirical Scholarship*. Emerald Group Publishing Limited.

Oyer, P. (2000). A theory of sales quotas with limited liability and rent sharing. *Journal of Labor Economics*, 18:405–426.

Parrinnello, R. (2017). Using accelerators to drive winning sales teams. *Workspan*, (August):32–39.

Pierce, L., Rees-Jones, A., and Blank, C. (2020). The negative consequences of loss-framed performance incentives. *NBER working paper*, (26619).

Royer, H., Stehr, M., and Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*, 7(3):51–84.

Saez, E. (2010). Do taxpayers bunch at kink points? *American economic Journal: economic policy*, 2(3):180–212.

Scudder, M. D. (2017). Will the dol rescind the tip pool rule? *The National Law Review*.

Seibold, A. (2021). Reference points for retirement behavior: Evidence from german pension discontinuities. *American Economic Review*, page forthcoming.

Spence, M. (1977). Nonlinear prices and welfare. *Journal of Public Economics*, 8:1–18.

Venkatesh, V. and Agarwal, R. (2006). Turning visitors into customers: A usability-centric perspective on purchase behavior in electronic channels. *Management Science*, 52(3):367–382.

A Appendix A: Theory

This Appendix derives the implications of three models for the distribution of output in the neighborhood of the two types of thresholds in Firm A's sales compensation scheme– the *kinks* (1.3T and 1.6T) at which only the commission rate increases, and the *target* (T) where an increase in the commission rate is combined with a small bonus.[45] The **standard model** assumes workers derive utility only from money and leisure, and have quasi-linear utility in money. In the **symbolic rewards (SR)** model, workers perceive attaining the target or kink as a (lump sum) psychic reward that directly raises their utility. In the **reference point (RP)** model, the target or kink creates a reference point, and induces loss aversion relative to it. We begin by formally stating the assumptions of each model, then derive the three models' predictions for the kinks. We conclude by deriving the predictions for the target. All the predictions are summarized in Table 2 of the paper.

A.1 Assumptions

All three models are special cases of the following framework. The worker's utility is:

$$u(y, q) = v(y) - \frac{c(q)}{\psi}, \tag{1}$$

where y is income and q is output. Marginal costs of producing output, $\frac{c'(q)}{\psi}$ are increasing, convex and differentiable everywhere; $\psi > 0$ represents cross-sectional heterogeneity in these costs. We will refer to workers with higher levels of ψ (lower effort costs) as higher-ability workers. Utility derived from income, or *effective income*, is given by $v(y)$. Working in terms of effective income allows us to express symbolic rewards and loss aversion as equivalent to the following changes to the commission rate, relative to a threshold output level, y^t:

- the lump-sum symbolic award is equivalent to a cash bonus at the threshold.
- loss aversion is equivalent to a convex kink in the commission schedule at the threshold.

In more detail, the **standard model** assumes that:

$$v(y) = y \tag{2}$$

The **symbolic rewards** model assumes that:

$$v(y) = \begin{cases} y & \text{if } y < y^t \\ y + B & \text{if } y \geq y^t \end{cases} \tag{3}$$

[45] All the models in this appendix are derived for the case of a single decisionmaker. Thus they abstract from issues of free-riding and co-ordination that might arise within Firm A's sales teams.

where y^t is the worker's income at the threshold and $B > 0$.

Finally, the **reference point** model assumes that:

$$v(y) = \begin{cases} y + \theta \cdot (y - y^t) & \text{if } y < y^t \\ y + \theta \cdot (y - y^t) & \text{if } y \geq y^t \end{cases} \tag{4}$$

where $\theta \in [0, 1]$ indicates the strength of reference-dependent preferences.[46]

Notably, all our models ignore the 'lumpy' and uncertain nature of sales at Firm A, which significantly constrain a team's ability to exactly hit the sales bin that contains a salient threshold. Thus we view our predictions as extreme versions of what we expect to observe, which is excess or missing mass in bins that contain a threshold, with offsetting changes in density in output regions that are either below or above the threshold's.

A.2 Predictions for the kinks (1.3T and 1.6T)

For all three models, this Section compares the predicted behavior of workers under two conditions. In the first condition, workers face a linear baseline commission rate:

$$y = y^K + w_1(q - q^K), \forall q \tag{5}$$

In the second, we add an accelerator to this schedule, as follows:

$$y = \begin{cases} y^K + w_1(q - q^K) & \text{if } q < q^K \\ y^K + w_2(q - q^K) & \text{if } q \geq q^K \end{cases} \tag{6}$$

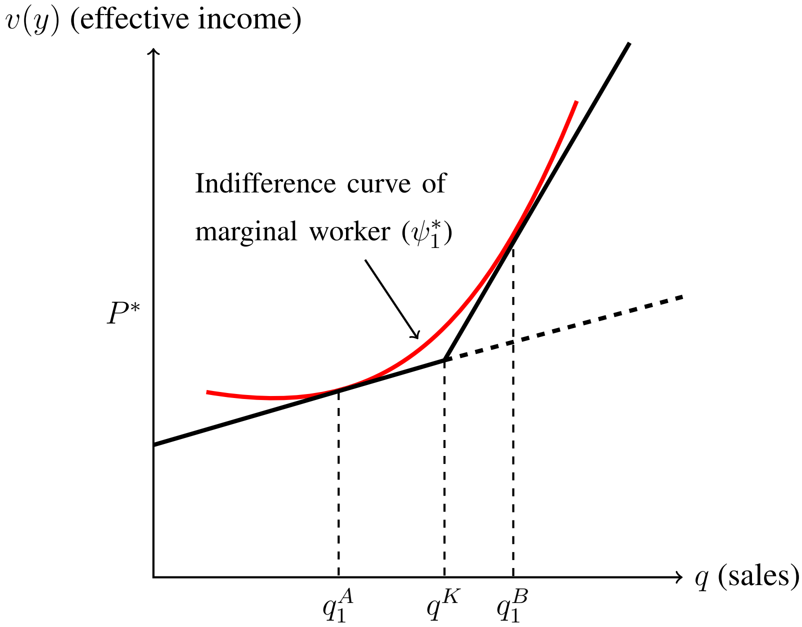
where $w_1 < w_2$, and y^K and q^K are the income and output levels at the kink in the budget constraint.

For the baseline (unkinked) budget constraint in the standard model, there is a one-to-one mapping between the ability parameter, ψ, and the worker's optimal sales output, q. If the density of ability, $f(\psi)$ is strictly positive and continuous on $(0, \infty)$, this yields a continuous density of observed output levels. Following Saez (2010), we can then derive predictions for the effect of introducing a concave kink at q^k on the distribution of output levels.

[46] In order to facilitate comparisons between situations with and without a reference point, our one-parameter reference point model departs from Kőszegi and Rabin (2006)'s (in which $v(y) = y + \eta\lambda(y - y^t)$ if $y < y^t$, $v(y) = y + \eta(y - y^t)$ if $y \geq y^t$, $\eta \in [0, 1]$ and $\lambda > 1$) If effective income in the absence of a reference point is given by $v(y) = y$, then Koszegi and Rabin's η parameter makes workers care more about income on both sides of an introduced reference point, which would raise overall labor supply even if there was no kink in the utility function (i.e. even if $\lambda = 0$). In our version, introducing a reference point raises marginal utility below the reference point, and reduces marginal utility above the reference point, by equal amounts. In this sense we preserve the 'overall' marginal utility of income when we introduce a reference point, in contrast to Koszegi-Rabin's approach which would raise 'overall' marginal utility whenever $\eta > 0$.

A.2.1 The standard model

In the standard model, $v(y) = y$. In effective income terms, the budget constraint and the indifference curve of a marginal worker (who is indifferent between the two segments of the budget constraint) are shown below:



In this model, workers with ability (ψ), below ψ_1^* will sell less than q_1^A in the absence of an accelerator, and will not change their output (q) when the accelerator is introduced.[47] Workers with ability above ψ_1^* will raise their (q) to a level above q_1^B. Finally, no workers will locate between q_1^A and q_1^B.

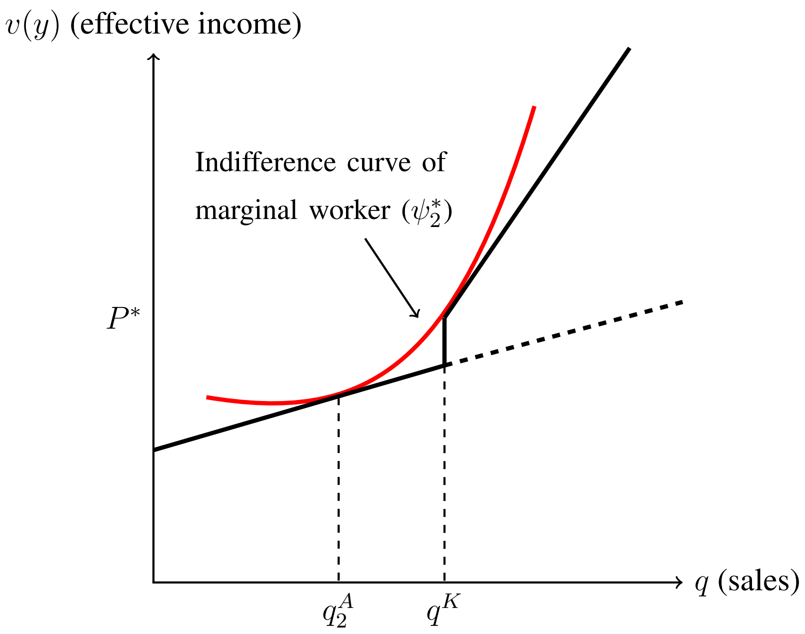
Adding an accelerator therefore:

- leaves the density of sales to the left of q_1^A unchanged;
- reduces the density of sales (to zero) in a region that (strictly) includes the kink (between q_1^A and q_1^B);
- raises the number of workers producing strictly more than the kink (q^K);
- no mass point (bunching) is induced.

A.2.2 The symbolic rewards model

In the symbolic rewards model, $v(y) = y$ if $y < y^K$, and $v(y) = y + B$ if $y \geq y^K$. Now, the effective-income budget constraint and marginal indifference curve take the following form:

[47] ψ_1^* is defined as the ability of the marginal worker, who is indifferent between q^A and q^B after the accelerator is introduced.



Now, workers with ability, ψ, below ψ_2^* will sell less than q_2^A in the absence of an accelerator, and will not change their output (q) when the accelerator is introduced. Because the symbolic reward makes it more attractive to reach the kink, $q_2^A < q^A$. Some of the workers with ability, ψ, above ψ_2* will raise their output (q) to exactly q^K. These are the workers with ability in some interval (ψ_2^*, ψ_2'), where ψ_2' corresponds to tangency (from the right) at q^K.[48]

The remaining workers (with $\psi > \psi_2'$) will raise their output levels to a point strictly above q^K. No workers will locate between q_2^A and q^K.

Adding an accelerator therefore:

- leaves the density of sales to the left of q_2^A unchanged;
- reduces the density of sales (to zero) *below* the kink (between q_2^A and q^K);
- raises the number of workers producing strictly more than the kink (q^K);
- creates a mass point (bunching) at q^K.

A.2.3 The reference point model

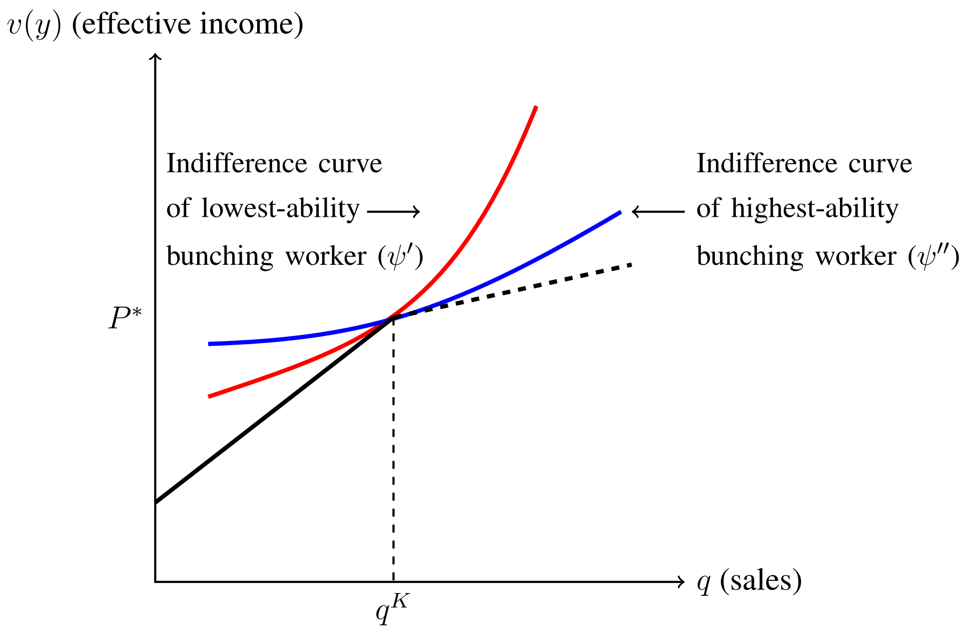
In this model, $v(y) = y + \theta(y - y^K)$ if $y < y^K$, and $v(y) = y - \theta(y - y^K)$ if $y \geq y^K$. Here, the effects of introducing an accelerator depend on the relative size of the loss-aversion parameter (θ) and the jump in the commission rate ($\frac{w_1}{w_2}$) at q^K. There are two cases:

1. *Weak Loss Aversion* ($\frac{1+\theta}{1-\theta} < \frac{w_1}{w_2}$):
When loss aversion is weak relative to the marginal wage increase at the kink, the marginal

[48]The indifference curve of the marginal worker will in general be steeper than the budget constraint at q^K; with probability measure zero it will be tangent there, in which case there is no bunching.

effective-income gradient to the left of the kink ($\frac{dv}{dq}$) remains lower than the gradient to the right of the kink. Thus, in effective income terms, the worker still faces a concave kink at q^K, so the predictions are (qualitatively) identical to the staandard model: the accelerator should create a gap in the density of sales around (i.e. on both sides of) the kink point. No bunching of sales should be observed.

2. *Strong Loss Aversion* ($\frac{1+\theta}{1-\theta} > \frac{w_1}{w_2}$):
   Now the effective-income gradient falls at the kink, creating a convex kink (in effective income terms):



Now, *all* workers who previously sold less than q^K will increase their output when the accelerator is introduced, because the marginal psychic rewards to working have increased from w_1 to $w_1(1+\theta)$. Conversely, *all* workers who previously sold more than q^K will reduce their output when the accelerator is introduced, because the marginal psychic rewards to working have decreased from w_2 to $w_2(1-\theta)$. Some of both groups of workers will move exactly to the kink point, q^K.

Adding an accelerator therefore:

- creates a mass point (bunching) at q^K;
- reduces the number of workers producing strictly less than q^K;
- *reduces* the number of workers producing strictly more than q^K.

Combining cases 1 and 2, we can summarize the predictions of the reference point model as follows: In the only circumstances where loss aversion leads to bunching at the kink, adding an accelerator is predicted to reduce the number of workers producing strictly more than the kink. This is not consistent with our empirical results.

A.3 Predictions for the Target (T)

With the preceding tools in hand, it is straightforward to derive the three models' predictions for the output distribution around the target, T, where an accelerator is combined with a cash bonus. To see this, note first that the standard model is now formally equivalent to the symbolic rewards (SR) model, with the cash bonus now playing the same role as the SR model's *psychic* bonus. Thus, we expect bunching at the target and excess mass to its right. By the same reasoning, the qualitative predictions of the symbolic rewards model are the same at the target as they were at the kinks– the only difference between the target and kinks is that the bonus at the target includes a cash component in addition to the symbolic one. Finally, the presence of a cash bonus means that, in general, the reference point (RP) model –like the standard and SR models– *also* predicts some bunching at the target. Extending previous reasoning, this bunching will be accompanied by excess mass to the right of the target when loss aversion is weak, and missing mass if loss aversion is strong.

B Appendix B: Detailed Methods

B.1 Statistical Significance of Bunching in the Monthly Sales Distributions

To estimate the significance of bunching in Figure 4(b), we estimate a regression of the following form:

$$c_j = \sum_{i=0}^{p} \beta_i \cdot (z_j)^i + \sum_{i=1,1.3,1.6} \gamma_i \cdot \mathbb{1}[z_j = i] + \epsilon_j,$$

where c_j is the number of store-month observations in bin j, z_j is the level of relative performance in bin j ($z_j = 0.1, 0.2, ..., 4$), and p is the order of polynomial. Then we use the predicted values from the above regression to construct the counterfactual bin counts, i.e. $\hat{c_j} = \sum_{i=0}^{p} \hat{\beta_i} \cdot (z_j)^i$. The bunching is then estimated as the excess mass relative to the counterfactual bin counts, i.e. $b_j = c_j - \hat{c_j} = \hat{\gamma_j}$, where j=1, 1.3 and 1.6 in our context. The solid curve in Figure 4(b) shows the counterfactual distribution with an eighth-degree polynomial.

B.2 Calculating Baseline Densities in the Output Bins that 'Typically' Contain the Target and Kinks

Here we describe calculate the baseline densities in the output that 'typically' contain the target and two kinks on control days. These are used in Section 5.1 to assess the magnitude of our main estimated coefficients.

We start by computing the mean density in a target's or kink's bin at each store i, separately for T, 1.3T, and 1.6T. For instance, the vector for T at store i, $\boldsymbol{\omega_i^T}$, reflects the likelihood that each bin contains T, i.e. $\boldsymbol{\omega_i^T} = (\omega_{i,1}^T, ..., \omega_{i,N_i}^T)$, where N_i is the number of bins for store i. Then, we restrict to the control sample to compute the mean density of every bin in which store i's sales output falls in, and call it $\boldsymbol{f_i} = (f_{i,1}, ..., f_{i,N_i})$. To compute the likelihood that store i's sales output falls in the bin that contains T, we compute the mean of $f_{i,b}$ across its bins, weighted by $\omega_{i,b}$, i.e. $F_i^T = \boldsymbol{\omega_i^T} \times \boldsymbol{f_i'}$. Thus, F_i^T reflects the likelihood that store i's output falls in the bins where T typically occurs. Finally, we take the sum of F_i^T across all stores to get the baseline density, weighted by the number of observations contributed by each store in the overall sample. This baseline density is estimated to be 10.3 percentage points for the target T. In other words on days when *no* target or kink is attainable, the mix of bins in which the target would typically fall contain 10.3$\%$ of the density of store sales. The baseline densities for the two pure kinks 1.3T and 1.6T are defined analogously and equal 7.7 percentage points for 1.3T, and 5.9 percentage points for 1.6T.

C Appendix C: Robustness

C.1 Defining Attainability

Table C1: Robustness Check – Attainability

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month		
	T	$1.3T$	$1.6T$
$K_{imt,-5}$	-0.0266	-0.0017	-0.0497***
	(0.0205)	(0.0459)	(0.0120)
$K_{imt,-4}$	-0.0245	-0.0410**	0.0242
	(0.0233)	(0.0182)	(0.0388)
$K_{imt,-3}$	0.0293	-0.0247	-0.0022
	(0.0366)	(0.0319)	(0.0226)
$K_{imt,-2}$	-0.0223	0.0137	-0.0191
	(0.0257)	(0.0341)	(0.0448)
$K_{imt,-1}$	-0.0306	-0.0028	-0.0231
	(0.0208)	(0.0257)	(0.0245)
$\boldsymbol{K_{imt,0}}$	**0.0811****	**0.1028****	**0.1312****
	(0.0330)	**(0.0421)**	**(0.0580)**
$K_{imt,1}$	0.0331	-0.0194	-0.0192
	(0.0273)	(0.0192)	(0.0226)
$K_{imt,2}$	0.0286	-0.0238	0.0094
	(0.0259)	(0.0157)	(0.0347)
$K_{imt,3}$	-0.0009	-0.0112	0.0003
	(0.0143)	(0.0161)	(0.0213)
$K_{imt,4}$	0.0327	0.0202	0.0310
	(0.0206)	(0.0225)	(0.0319)
$K_{imt,5}$	-0.0023	0.0051	-0.0185***
	(0.0111)	(0.0170)	(0.0043)
N	2697816	2693899	2691619

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
This Table replicates columns (4)-(6) of Table 3 using a different definition of attainability. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

Table C1 defines a threshold as attainable if it is within three times the store's mean daily sales during the previous same *days of the week* in the current month. The results are very simlar to our main analysis.

C.2 Alternative Control Groups

In our main analysis, the control group included all non-last days in a month, plus non-last days on which no threshold was attainable. However, it is possible that a forward-looking team can strategically manipulate the presence or the position of an attainable threshold on the last day of a month. To alleviate this concern, Table C2 excludes the week preceding the month's last day from the control group, since it becomes easier to manipulate the position to a threshold when it proceeds near the end of a month. Another possibility is that last days are so unique that we cannot use non-last days to control for their sales cycle. In Table C3, we use only last days when no target or kink is attainable as controls. Due to the much smaller sample size, we control for Bin$\times$Store and Bin$\times$Holiday fixed effects in columns (4)-(6). In both cases, the estimated effects are also very similar to our main results.

Table C2: Excluding the Week Preceding the Month's Final Day from the Control Group Team Output Distributions (Replicating Table 3)

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month					
	Specification 1: Raw Estimates			Specification 2: FEs Controlled		
	T	$1.3T$	$1.6T$	T	$1.3T$	$1.6T$
$K_{imt,-5}$	0.0542	0.0351	-0.0125***	-0.0158	-0.0356	-0.0551***
	(0.0327)	(0.0333)	(0.0012)	(0.0283)	(0.0344)	(0.0109)
$K_{imt,-4}$	0.0280	0.0487	0.0909	-0.0449*	-0.0315	0.0542
	(0.0227)	(0.0341)	(0.0566)	(0.0249)	(0.0260)	(0.0519)
$K_{imt,-3}$	0.1404***	0.0441	0.0208	0.0505	-0.0255	-0.0074
	(0.0416)	(0.0316)	(0.0321)	(0.0399)	(0.0343)	(0.0251)
$K_{imt,-2}$	0.0612**	0.0968**	0.0160	-0.0284	0.0054	-0.0580
	(0.0302)	(0.0381)	(0.0285)	(0.0282)	(0.0340)	(0.0387)
$K_{imt,-1}$	0.0351*	0.0956***	0.0388	-0.0431**	0.0329	-0.0263
	(0.0208)	(0.0349)	(0.0352)	(0.0191)	(0.0313)	(0.0263)
$\boldsymbol{K_{imt,0}}$	**0.1730*** **	**0.1727*** **	**0.1875*** **	**0.0940*** **	**0.0896** **	**0.1416** **
	(0.0343)	**(0.0408)**	**(0.0701)**	**(0.0337)**	**(0.0427)**	**(0.0622)**
$K_{imt,1}$	0.0843***	0.0245	0.0125	0.0328	-0.0204	-0.0200
	(0.0254)	(0.0209)	(0.0243)	(0.0264)	(0.0210)	(0.0250)
$K_{imt,2}$	0.0681**	0.0122	0.0375	0.0260	-0.0245	0.0140
	(0.0275)	(0.0171)	(0.0353)	(0.0262)	(0.0162)	(0.0363)
$K_{imt,3}$	0.0278	0.0122	0.0125	0.0003	-0.0079	0.0041
	(0.0169)	(0.0169)	(0.0242)	(0.0146)	(0.0171)	(0.0217)
$K_{imt,4}$	0.0520**	0.0369	0.0375	0.0326	0.0224	0.0318
	(0.0219)	(0.0233)	(0.0334)	(0.0212)	(0.0230)	(0.0338)
$K_{imt,5}$	0.0036	0.0122	-0.0125***	-0.0024	0.0053	-0.0143***
	(0.0115)	(0.0173)	(0.0012)	(0.0114)	(0.0181)	(0.0045)
N	2150998	2146959	2145790	2150998	2146959	2145790

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
This Table replicates Table 3, excluding the last seven days of every month from the control group. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

Table C3: Using Only Last Days as Control Days (Replicating Table 3)

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month					
	Specification 1: Raw Estimates			Specification 2: FEs Controlled		
	T	$1.3T$	$1.6T$	T	$1.3T$	$1.6T$
$K_{imt,-5}$	0.0549*	0.0354	-0.0122***	-0.0056	-0.0263	-0.0295***
	(0.0328)	(0.0333)	(0.0012)	(0.0303)	(0.0352)	(0.0108)
$K_{imt,-4}$	0.0288	0.0490	0.0912	-0.0530*	-0.0044	0.0176
	(0.0228)	(0.0341)	(0.0567)	(0.0272)	(0.0300)	(0.0608)
$K_{imt,-3}$	0.1412***	0.0443	0.0211	0.0721*	-0.0289	-0.0052
	(0.0416)	(0.0316)	(0.0321)	(0.0399)	(0.0361)	(0.0306)
$K_{imt,-2}$	0.0620**	0.0971**	0.0163	-0.0047	0.0208	-0.0379
	(0.0302)	(0.0382)	(0.0285)	(0.0271)	(0.0387)	(0.0359)
$K_{imt,-1}$	0.0359*	0.0958***	0.0390	-0.0185	0.0297	-0.0334
	(0.0209)	(0.0350)	(0.0352)	(0.0196)	(0.0323)	(0.0291)
$\boldsymbol{K_{imt,0}}$	**0.1738***	**0.1729***	**0.1878***	**0.1002***	**0.1309***	**0.1343**
	(0.0342)	**(0.0407)**	**(0.0702)**	**(0.0373)**	**(0.0423)**	**(0.0665)**
$K_{imt,1}$	0.0851***	0.0248	0.0128	0.0271	-0.0306	-0.0026
	(0.0254)	(0.0210)	(0.0243)	(0.0273)	(0.0226)	(0.0231)
$K_{imt,2}$	0.0689**	0.0124	0.0378	0.0405	-0.0183	0.0233
	(0.0275)	(0.0171)	(0.0353)	(0.0279)	(0.0156)	(0.0339)
$K_{imt,3}$	0.0286*	0.0124	0.0128	-0.0053	-0.0204	0.0101
	(0.0170)	(0.0169)	(0.0242)	(0.0166)	(0.0201)	(0.0242)
$K_{imt,4}$	0.0528**	0.0371	0.0378	0.0365	0.0304	0.0294
	(0.0219)	(0.0234)	(0.0334)	(0.0221)	(0.0217)	(0.0330)
$K_{imt,5}$	0.0044	0.0124	-0.0122***	-0.0082	-0.0001	-0.0277***
	(0.0115)	(0.0173)	(0.0012)	(0.0128)	(0.0189)	(0.0085)
N	80506	76467	75298	80506	76467	75298

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
This Table replicates columns Table 3, using only last days of the month as the control group. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

D Appendix D: Heterogeneity

Table D1: Heterogeneity Examination – Turnover Events

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month			
	T	1.3*T* & 1.6*T*	*T*	1.3*T* & 1.6*T*
$K_{imt,0}$	0.0814**	0.1086***	0.0771**	0.1111**
	(0.0352)	(0.0406)	(0.0363)	(0.0425)
$K_{imt,0}\times$ Turnover$_{im}$	0.0861	0.1267		
	(0.1033)	(0.1086)		
$K_{imt,0}\times$ Turnover$_{im2}$			0.0843	0.0782
			(0.0887)	(0.0905)
Bin×Store×DOW	Yes	Yes	Yes	Yes
Bin×Holiday	Yes	Yes	Yes	Yes
N	2697816	2697791		

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This Table replicates columns (4)-(6) of Table 3, interacting the threshold-containing bin with indicators for recent turnover on the sales team. Turnover$_{im}$ is an indicator variable, taking a value of 1 if there is turnover event during the current month m at store i. Turnover$_{im2}$ is an indicator variable, identifying month m if there is turnover event at store i, and identifying the month $m + 1$ following the turnover. To economize on degrees of freedom, the regressions for 1.3T and 1.6T are combined. Like columns (4)-(6) of Table 3, all regressions include variables identifying five bins on either side of the threshold. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

Table D2: Heterogeneity Examination – Team Size Effects

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	T	$1.3T$ & $1.6T$
$K_{imt,0}\times$ Size=2	-0.0866	0.0213
	(0.0758)	(0.0861)
$K_{imt,0}\times$ Size=3	0.0965**	0.1067**
	(0.0454)	(0.0456)
$K_{imt,0}\times$ Size≥4	0.1442***	0.2050**
	(0.0492)	(0.0801)
Bin×Store×DOW	Yes	Yes
Bin×Holiday	Yes	Yes
N	2778794	2778794

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
This Table replicates columns (4)-(6) of Table 3, interacting the threshold-containing bin with indicators for team size. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Like columns (4)-(6) of Table 3, all regressions include variables identifying five bins on either side of the threshold. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.

Table D3: Heterogeneity Examination – Firm Tenure Effects

	Dependent variable: Indicator for team sales falling in a given bin on the last day of the month	
	T	1.3*T* & 1.6*T*
$K_{imt,0}\times$ c.Tenure	0.0128**	0.0190**
	(0.0058)	(0.0090)
Bin×Store×DOW	Yes	Yes
Bin×Holiday	Yes	Yes
N	2778794	2778794

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This Table replicates columns (4)-(6) of Table 3, interacting the threshold-containing bin with a continuous measure of the team's mean tenure with Firm A. To economize on degrees of freedom, the regressions for $1.3T$ and $1.6T$ are combined. Like columns (4)-(6) of Table 3, all regressions include variables identifying five bins on either side of the threshold. Standard errors reported in parentheses are clustered at the store level. Please see notes to Table 3 for additional details.